VOL. 48 1957

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

BOARD OF EDITORS*

Wm. Clark Trow, *Chairman*
University of Michigan

Stephen M. Corey
Teachers College
Columbia University

J. B. Stroud
State University of Iowa

*Appointed by the American Psychological Association

EDITORIAL CONSULTANTS

Paul Blommers
Chester W. Harris
Harold E. Jones

H. H. Remmers
Percival M. Symonds

Miles A. Tinker
Alexander G. Wesman
Paul A. Witty

American Psychological Association

Bureau Ednl. Psy. Research
DAVID HA[illegible]ING COLLEGE
Dated ....
Accs. No ....

CONTENTS OF VOLUME 48

Aiken, L. R., Jr. See Dreger, R. M.

Allison, Louise. See Greenberg, H. M.

Ausubel, D. P., Robbins, Lillian Cukier, and Blake, E., Jr. Retroactive Inhibition and Facilitation in the Learning of School Materials 334

Ausubel, D. P., Schpoont, S. H., and Cukier, Lillian. The Influence of Intention on the Retention of School Materials 87

Barch, A. M. The Relation of Departure Time and Retention to Academic Achievement 352

Beilin, H., and Werner, Emmy. Sex Differences Among Teachers in the Use of the Criteria of Adjustment 426

Blake, E., Jr. See Ausubel, D. P.

Blommers, P. See Stroud, J. B.

Brown, M. H., and Bryan, G. Elizabeth. Sex as a Variable in Intelligence Test Performance 273

Bryan, G. Elizabeth. See Brown, M. H.

Calvin, A. D., McGuigan, F. J., and Sullivan, M. W. A Further Investigation of the Relationship Between Anxiety and Classroom Examination Performance 240

Cankardas, A. See Wysocki, B. A.

Cantril, H. See Toch, H.

Carrier, N. A. The Relationship of Certain Personality Measures to Examination Performance Under Stress 510

Clark, E. L. Item Difficulties Based on End Segments 457

Colville, Frances M. The Learning of Motor Skills as Influenced by Knowledge of Mechanical Principles 321

Cook, D. A Comparison of Reading Comprehension Scores Obtained Before and After a Time Announcement 440

Cukier, Lillian. See Ausubel, D. P.

Della-Piana, G. M. Searching Orientation and Concept Learning 245

Dennis, W. Animistic Thinking among College and High School Students in the Near East 193

Drasgow, J. Problems of Progeny Related to Paternal Education 521

Dreger, R. M., and Aiken, L. R., Jr. The Identification of Number Anxiety in a College Population 344

Eglash, A. Changes in Opinionation During a Psychology Course 164

Egner, R. E., and Obelsky, A. J. Effect of Stereotyped Attitudes on Learning 207

English, H. B. Chronological Divisions of the Life Span 437

Fast, Irene. Kindergarten Training and Grade I Reading 52

Faw, V. Learning to Deal with Stress Situations 135

Ferguson, Eva D. An Evaluation of Two Types of Kindergarten Attendance Programs 287

Fewell, Mildred. See Greenberg, H. M.

Finley, Carmen J. See Holmes, J. A.

Fishman, J. A. The MTAI in an American Minority-Group Setting. I. Differences Between Test Characteristics for Norm and Non-norm Populations 41

French, J. W. The Factorial Invariance of Pure-Factor Tests . . . 93
Furness, Edna L. Listening: A Case of Terminologiial Confusion . . 477
Gage, N. L. Logical Versus Empirical Scoring Keys: The Case of the MTAI 213
Goertzen, S. M. A Study of Teachers' and Psychologists' Ability to Predict Seventh Graders' Opinions of Certain Behaviors of Their Peer Group 166
Greenberg, H. M., Allison, Louise, Fewell, Mildred, and Rich, C. The Personality of Junior High and High School Students Attending a Residential School for the Blind 406
Heber, Mary E. See Heber, R. F.
Heber, R. F., and Heber, Mary E. The Effect of Group Failure and Success on Social Status 129
Hill, L. B. A Second Quarter Century of Delayed Recall or Relearning at 80 65
Holmes, J. A., and Finley, C. J. Under- and Over-Age Grade-Placements and School Achievement 447
Holmes, J. A., and Hyman, W. Spelling Disability Asyntaxia in a Case Involving Injury to the Language Formulation Area of the Brain 542
Hountras, P. T. The Relationship between Pre-Admission Data and Achievement of Foreign Graduate Students 157
Hyman, W. See Holmes, J. A.
Kittell, J. E. An Experimental Study of the Effects of External Direction During Learning on Transfer and Retention of Principles 391
Lauber, Margaret. See Stroud, J. B.
McDonald, A. S. Influence of a College Improvement Program on Academic Performance 171
McGuigan, F. J. See Calvin, A. D.
McKeachie, W. J., and Solomon, D. Retention of General Psychology . . 110
Macomber, F. G. See Siegel, L.
Massey, H. W. See Vineyard, E. E.
Mayer, Sylvia R. The Effect of Induced Tension During Training of Visual Form Recognition 11
Mayo, G. D. Differentiating Characteristics of a Group of Students Having Psychological Problems 359
Medley, D. M. See Mitzel, H. E.
Metaweh, I. E. Recruiting Prospective Rural Teachers at the Men's Teachers' College, Tripoli, Libya 69
Mitzel, H. E., and Medley, D. M. Pupil Growth in Reading—An Index of Effective Teaching 227
Newman, S. E. Student vs. Instructor Design of Study Method 328
Obelsky, A. J. See Egner, R. E.
Parsons, T. S. A Comparison of Instruction by Kinescope, Correspondence Study and Customary Classroom Procedures 27
Rich, C. See Greenberg, H. M.
Robbins, Lillian Cukier. See Ausubel, D. P.
Rosenthal, E. Some Relationships Between Sociometric Positions and Language Structure of Young Children 483

Scarf, R. C. Differential Scores and Unstable Personality 278
Scarborough, B. B. See Wright, J. C.
Schpoont, S. H. See Ausubel, D. P.
Schwartz, M. Transfer of Reading Training From Nontechnical to Technical Material 498
Siegel, L., and Macomber, F. G. Comparative Effectiveness of Televised and Large Classes and of Small Sections 371
Silberman, H. F. Effects of Praise and Reproof on Reading Growth in a Non-Laboratory Classroom Setting 199
Simpson, R. H. A Procedure for Managing Teaching-Learning Situations in Educational Psychology 411
Solomon, D. See McKeachie, W. J.
Stroud, J. B. The Intelligence Test in School Use: Some Persistent Issues 77
Stroud, J. B., Blommers, P., and Lauber, Margaret. Correlation Analysis of WISC and Achievement Tests 18
Sullivan, M. W. See Calvin, A. D.
Taylor, M. Predicting Interests From Different Kinds of Cues 534
Toch, H., and Cantril, H. A Preliminary Inquiry into the Learning of Values 145
Tyler, F. T. Stability of Intra-Individual Patterning of Measures of Adjustment during Adolescence 217
Ullmann, C. A. Teachers, Peers and Tests as Predictors of Adjustment 257
Vineyard, E. E., and Massey, H. W. The Interrelationships of Certain Linguistic Skills and Their Relationship with Scholastic Achievement When Intelligence is Ruled Constant 279
Werner, Emmy. See Beilin, H.
Wertheimer, Rita R. Consistency of Sociometric Status Position in Male and Female High School Students 385
Whittington, R. W. T. The Assessment of Potential Musical Ability in Secondary School Children 1
Worcester, D. A. Learning Ability and Retention After Long Periods .. 505
Wright, J. C., and Scarborough, B. B. The Interrelationship of Area Test Scores and Cooperative General Culture Test Scores 460
Wysocki, B. A. Assessment of Intelligence Level by the Rorschach Test as Compared with Objective Tests 113
Wysocki, B. A., and Cankardas, A. A New Estimate of Polish Intelligence 525

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 48 January, 1957 Number 1

THE ASSESSMENT OF POTENTIAL MUSICAL ABILITY IN SECONDARY SCHOOL CHILDREN

R. W. T. WHITTINGTON

(Former Music Master, Otago Boys' High School, Dunedin, New Zealand

In order to ascertain the degree of musical ability in all first-year boys it was decided to carry out tests using material derived from similar investigations by Dr. C. H. Wing of London (*16*). By applying this material to two carefully selected groups it was hoped, *inter alia*, to evolve a shorter battery of tests which could be used to ascertain musical ability.

While seeking a general approach to the question the author consulted past work in the field of musical ability, and on reading James L. Mursell's *The Psychology of Music*, came across a statement that formed the basic approach to his problem: "We must try," wrote Mursell, "our developed tests upon individuals known to be conspicuously musical and those known to be conspicuously nonmusical to try to discover where the most crucial and significant performances are located. Strange to say," he added, "this obvious procedure has only been adopted by Stumf and by Ravecz—and in each case only a single individual subject was involved" (*10*). Dr. Wing used a battery of seven tests which included tests of perceptual efficiency and tests of musical preference. In some of his tests Wing's approach was that of the Gestalt psychologists—a school of psychology which emphasized the perception of patterns, shapes, and tunes as wholes. The primary object of Wing's research was "to examine the hypothesis of a general factor for musical ability, and to explore the influence of more specialized factors such as might obscure this general factor or interfere with its measurement." Wing found that there was no great difference as

Bureau Ednl. Research
DAVID COLLEGE
Dated

regards the efficiency of the individual tests, but found the phrasing test which had a saturation coefficient of nearly 0.80 the best.

In order to find the most significant differences in performance it was proposed to follow up Mursell's premise and to try Wing's tests upon individuals known to be conspicuously musical and those known to be conspicuously nonmusical. From this it was hoped to reduce Wing's battery of seven tests to a more economical number—if possible to two or three. Most important, then, was the selection of the two groups. Mursell's use of the phrase "conspicuously musical" made it imperative to select only those pupils who had proved themselves both in the theory and practice of music. Ages in this group ranged from thirteen years eleven months to eighteen years six months, the mean age being sixteen years three months. My idea was to select pupils from the sixteen to eighteen age group, for by doing so I was certain of obtaining only those who had advanced music certificates. However, four pupils under fifteen years of age were included in the group. I knew two of these pupils to be conspicuously musical, and information concerning the other two was given by their respective college music directors. Two pupils possessed Associate Diplomas of recognized music colleges, while two possessed the Licentiate. The remainder of the group had reached the higher grades and had gained either honours or merit in their theory and practical examinations. In all but four cases there was musical activity in the home. All twenty-four of the musical group were active participants in the musical life of their respective schools—orchestras, bands, choral clubs, and so on. Five of the group had shown promise in the field of composition and had actually had their works performed in public.

The second group was more difficult to select. To follow Mursell's premise faithfully this group had to be conspicuously nonmusical. Selection was based mainly on: first, complete dislike for music, and second, lack of musicality in parents, brothers, and sisters. For a final selection I used Wing's questionnaire. Subjects were asked to place a tick against the word that most nearly described their general attitude towards music: (a) very interested; (b) interested; (c) indifferent; (d) dislike.

They could add a plus or minus sign if they desired finer shades than those given. Secondly, the following was given to elicit information about their own musical activities and those of their parents, brothers, and sisters: Does anyone at home play an instrument? If so, state who they are, what they play, and roughly how much, e.g., "father, oboe, frequently." They were asked to use terms "frequently," "occasionally," or "seldom" to describe the amount of playing activity. For my final nonmusical group I selected those pupils who: (a) placed a tick against dislike; (b) recorded a "nil" answer to musical activity at home; (c) did not play an instrument and were not interested in learning to do so; (d) were not keen about school musical activities such as school singing (this was an individual question). Ages in this group ranged from fifteen years to seventeen years eleven months, the mean age being sixteen years six months.

Besides these two main groups I included a group of three pupils who, having had some years of music instruction, had given up because of lack of progress. I thought that, although a very small group, they might provide some further evidence to aid final analysis. For the experiment I would have liked a very much larger group, but the fact that testing had to be done out of school hours made it most difficult to collect a group enthusiastic enough to take the tests. As far as the musical group was concerned, directors of music of the secondary schools concerned sent only their best pupils. I considered this to be an excellent group.

To ascertain ability to appreciate certain aspects of musical material, the complete battery of Wing's tests was administered. This included chord analysis, pitch change, memory tests, rhythmic accent, harmony, intensity, and phrasing. While making preliminary arrangements for the investigation, approach had been made to various music teachers and others interested in music and their opinions sought on the nature of musical ability. Quite a number were emphatic about the possession of good intelligence. However, one particular teacher of pianoforte made this reply, "It is all a question of head and hand." Whether manual dexterity would prove to be a factor or not prompted me to include the Minnesota Rate of Manipulation Test. The first two individual tests were used—placing, and turning and placing. The first test involved the use of one hand, the second, two hands. To test intelligence Raven's Progressive Matrices, 1947, was used.

ANALYSIS OF THE DATA

Although sixty pupils were tested two groups each of twenty-four pupils were selected for final diagnosis. The first task was to ascertain significant differences between the musical scores of the two groups. The critical ratio proved to be 10.5. Similarly, critical ratios were obtained between the two groups on all musical tests:

Test 1	7.02
Test 2	8.78
Test 3	6.50
Test 4	4.27
Test 5	11.25
Test 6	4.64
Test 7	3.42

At the 0.01 level all the ratios were highly significant. The two groups were then regarded as distinct as far as their performances in the musical tests were concerned. The critical ratio between the percentile rankings on the Raven Intelligence, however, was 0.30. There was thus no outstanding difference between the two groups at the 0.01 level, and for the purpose of the investigation the two groups were regarded as being homogeneous as far as the matrices intelligence test was concerned.

The first part of the investigation was to follow up the Mursell premise. It was essential, of course, to deal separately with the two groups. By using factor analysis methods it was hoped that some factor or factors would be isolated which would indicate where the significant differences in performance lay. The next step, then, was to correlate the marks in the music tests and to factorize the correlations (Table I).

Only one factor was isolated from this group, a factor that accounted for some twenty-eight per cent of the performance. The rather high correlation of the Raven percentile suggested that intelligence as measured by this test was operating in some form or other. One was rather diffident to speculate just what the factor might be. The first factor saturations on tests 2 and 7 are high. Test 2 involved sensitivity to pitch relationships, and test 7 was one which required judgment of the more appropriate phrasing of a melody. Both tests involved a certain degree of appreciation of relationships, and one might say that test 7 involved

TABLE I—CORRELATION MATRIX AND FIRST FACTOR SOLUTION
(Nonmusic group)

Test	1	2	3	4	5	6	7	Raven	Σ
1	0.31	0.28	0.10	0.31	0.24	0.10	0.27	0.21	1.82
2	0.28	0.63	0.34	0.51	0.02	0.00	0.50	0.63	2.91
3	0.10	0.34	0.46	0.13	0.14	0.46	0.27	0.32	2.22
4	0.31	0.51	0.13	0.51	0.08	0.15	0.25	0.20	2.14
5	0.24	0.02	0.14	0.08	0.28	0.00	0.28	0.17	1.21
6	0.10	0.00	0.46	0.15	0.00	0.46	0.22	0.00	1.29
7	0.27	0.50	0.27	0.25	0.28	0.22	0.50	0.40	2.69
Raven	0.21	0.63	0.32	0.20	0.17	0.00	0.40	0.63	2.56
Σ	1.82	2.91	2.22	2.14	1.21	1.39	2.69	2.56	16.94
First factor loadings	0.44	0.71	0.54	0.52	0.29	0.34	0.65	0.62	

The mean communality of the first factor of this group was 0.282.

not only the ability to educe relations but also correlates. This is, of course, the idea underlying the matrices test. Test 3 had a saturation of 0.54. The test involved detection of a change of note in a short melodic phrase—a test which demanded retentivity.

It is not often an easy matter to listen first to a melodic phrase of a given number of notes, and then when the phrase is repeated to say which note is changed. However, the mere fact that test 2 involved a relationship of two notes—one higher or lower than the given note—and test 7 differentiated between staccato and legato playing may simply mean that both tests were measuring sensitivity. Whatever the factor operating through the performances of the nonmusical group, it would appear to be a significant element in the Raven test performance.

There appeared to be a factor which was responsible for some forty-four per cent of the performance in the musical group (Table II). The Raven loading was not as high in this group as it was in the case of the nonmusical group. Something was therefore operating through the performances of the musical group that was over and above intelligence as measured by the matrices test. Test 4 (judging the better rhythmic accent) and test 5 (judging the more appropriate of the two harmonizations) were the most outstanding performances. As all the first factor loadings were high, one may conclude a common factor running through

TABLE II—CORRELATION MATRIX AND FIRST FACTOR SOLUTION
(Musical group)

Test	1	2	3	4	5	6	7	Raven	Σ
1	0.66	0.13	0.66	0.57	0.39	0.39	0.46	0.36	3.62
2	0.13	0.54	0.17	0.54	0.47	0.37	0.44	0.18	2.84
3	0.66	0.17	0.66	0.64	0.53	0.31	0.16	0.42	3.55
4	0.57	0.54	0.64	0.65	0.65	0.46	0.40	0.40	4.31
5	0.39	0.47	0.53	0.46	0.65	0.37	0.57	0.47	4.10
6	0.39	0.37	0.31	0.40	0.37	0.52	0.41	0.52	3.35
7	0.46	0.44	0.16	0.40	0.37	0.42	0.57	0.20	3.21
Raven	0.36	0.18	0.42	0.40	0.47	0.52	0.20	0.52	3.07
Σ	3.62	2.84	3.55	4.31	4.10	3.35	3.21	3.07	28.05
First factor loadings	0.68	0.54	0.67	0.81	0.77	0.63	0.60	0.58	

The mean communality of the first factor of this group was 0.445.

all the tests. A residual matrix was obtained and McNemar's criterion applied.[1]

What, then, was this factor affecting the performances in all the tests? I hesitated to call it "musical ability." The group was a homogeneous one, and the critical ratios between the two groups was so highly significant that I could but regard the factor as one of "musical experience"—experience plus an intellectual element. This intellectual element I would not confuse with Spearman's "g." The experiential element would certainly account for the high saturation in test 5 (harmony), where ears trained in aural perception would easily detect those melodic lines in which chord progressions were obviously incorrect. The first factor loading in test 4 was 0.81. In this test Wing had included classical extracts which were obviously known by members of the musical group, for one had only to study their faces during the test performance to realise that tunes were familiar.

---

[1] McNemar's criterion: McNemar isolated factors until G_1 reached or fell below $\frac{1}{\sqrt{N}}$

Where $G_1 = G_s \div (1 - M_h^2)$

Where $G_s = SD$ of residuals after s factors

In the case of the residual Matrix for the musical group $G_1 = .19$ which fell below 0.20, i.e. $(1/\sqrt{24})$. Thus only one factor could be isolated from the performances of the musical group.

And so, to attempt an answer to Mursell's premise I could only suggest that the differences in performance lay in the fact that, in the case of the musical group, "experience" played a most important part.

It is worthy of note here that Wing obtained his three factors from an analysis of the results of his tests on a group of forty-three boys aged fourteen to sixteen. He chose this group because it was the largest one to which he had given the tests twice. His group was not, therefore, a homogeneous one. The mere fact that my nonmusical group lacked musical experience had, in all probability, accounted for low test scores. Worthy of note also was the fact that those subjects who took the tests but who were rejected because they had learned an instrument, or because of some form of musical expression in the home, scored a little higher than members of the nonmusical group. The critical ratios of the three girls who had had some three to four years of musical instruction were also significant and marked them off as a group whose performances in the music tests were much better than those of the nonmusical group.

The second part of the investigation involved the selection of a more economical battery of tests. Wing himself had stated that he did not object to a smaller number of tests. But to select such predictive tests from the original battery of seven it appeared that certain criteria should be satisfied. Tests should yield significant correlations with musical age in both the musical and the nonmusical groups; as the factor accounting for the best intercorrelations in the musical group appeared to involve an appreciative experiential element, the predictive tests should have comparatively low loadings in that factor; the intercorrelations among the predictive tests should be as low as possible.

Tests 2, 3, and 7 appeared to satisfy these criteria very well: they yielded significant and relatively constant correlations with musical age in both groups and their intercorrelations were low. When regression weights, determined by the Aitken-Thomson method of pivotal condensation, were applied, these three tests yielded most significant multiple correlations. Multiple correlation with the musical group was 0.92 and with the nonmusical group 0.90. Tests yielding very high correlations only in the musical group (tests 3, 4, and 5) when similarly weighted failed to give multiple correlations of any greater magnitude than the coefficients yielded by those tests in the two groups. This suggested that those particular tests were highly influenced by the "experiential factor" isolated in the musical group. It appeared, then, that as tests for selection purposes, tests 2, 3, and 4 were greatly to be preferred and appeared to be promising predictive instruments.

The manual dexterity tests failed to give any significant results. In the musical group the correlation of musical age and manual test 1 (placing) was 0.17, and with manual test 2 (turning and placing), 0.10. The correlations in the nonmusical group were lower still: 0.10, manual test 1, and −0.18 with manual test 2. No attempt was made to correlate the Raven percentiles with any of the manual dexterity tests. I had hoped that there would have been some significant correlation between musical age and manual test 2 which involved the use of both hands. Observance, however, indicated that facility at the piano in no way helped to turn and place the blocks quickly—it appeared to be even a hindrance. A pianist's hands and fingers are trained to a certain pattern, and I found that most of the good pianists of the musical group fumbled and dropped blocks a good deal. The only one who performed really well was a Chinese lad in the nonmusical group. As far as manual test 1 was concerned, I thought that the good violin players in the musical group might have performed well as this test called for dexterity in the bow hand. Such, however, was again not the case. It appeared, then, that manual dexterity could be ruled out as a factor to be considered as far as musical performance as measured by the Wing series of test was concerned.

SUMMARY

To assess musical ability and appreciation in the light of Mursell's premise and to select an economical battery of Wing's musical tests the following were administered: a) The Wing tests of musical intelligence, b) The 1947 Raven progressive matrices intelligence test (Australian Council of Educational Research), c) The Minnesota tests of manual dexterity (tests 1 and 2).

Tests were given to two carefully selected groups—musical, and nonmusical.

Factor analysis on the lines described by Cyril Burt and using the Thurstone centroid method of factor analysis was applied. The results seemed to indicate that the musical group was superior to the nonmusical group because of musical experience, an experience which accounted for some forty-four per cent of the performance. In the nonmusical group one was isolated which possibly utilized intelligence in a finer form to appreciate a "gestalt" quality.

Using the Aitken-Thomson method of pivotal condensation three tests yielded significant multiple correlations. These were tests 2, 3, and 7.

The correlation of performances in the musical tests with the matrices intelligence test was significant but not high.

Tests of manual dexterity were not to be regarded as playing any part in the diagnosis of musical ability as far as the Wing tests were concerned.

Further developments are possible using tests 2, 3 and 7 on untrained groups who wish to become musicians, with subsequent assessment of the predictive value of these tests when reliable measures of musical progress are available.

REFERENCES

1) A. S. Beckham, "A Study of the Social Background and Music Ability of Superior Negro Children," *Journal of Applied Psychology*, 26: 1942.

2) G. Brighouse and D. S. Ellis, "Effect of Music on Respiration and Heart Rate," *American Journal of Psychology*, 65: January, 1952.

3) H. Dewar, "A Comparison of Tests of Artistic Appreciation," *British Journal of Educational Psychology*, 8: 1951.

4) P. J. Fay and W. C. Middleton, "Relationships between Musical Talent and Preference for Different Types of Music," *American Journal of Educational Psychology*, 30: 1941.

5) R. Jamieson, "An Investigation into Songs Known by Scottish Children and Their Musical Preference," *British Journal of Educational Psycology*, 22: February, 1952.

6) C. J. Lamp and N. Keys, "Can Aptitude for Specific Musical Instruments be Predicted," *American Journal of Educational Psychology*, 26: 1935.

7) H. Levy and E. E. Preidel, *Elementary Statistics*. New York, Thomas Nelson & Sons, 1944.

8) H. Lowery, "Cadence and Phrase Tests in Music," *British Journal of Psychology*, 17: 1926.

9) J. Lundin, "A Preliminary Report on Some New Tests of Musical Ability," *Journal of Applied Psychology*, 28: 1944.

10) J. Mainwaring, "The Assessment of Musical Ability,"*British Journal of Educational Psychology*, 17: 1947. "Kinaesthetic Factors in the Recall of Musical Experience," *British Journal of Psychology*, 23: 1933. "Psychological Factors in the Teaching of Music," *British Journal of Educational Psychology*, 21: Part 2, 1951.

11) J. McLeish, "The Validation of Seashore Measures of Musical Talent by Factorial Methods," *British Journal of Psychology*, 3: Part 3, 1950.

12) J. L. Mursell, *The Psychology of Music*. New York, Norton, 1937.

13) Carl Seashore, *The Psychology of Musical Talent*. Silver, Burdett & Co., 1919.

14) B. Semeonoff, "A New Approach to the Testing of Musical Ability," *British Journal of Educational Psychology*, 30: 1941. "Further Developments in a New Approach to the Testing of Musical Ability," *British Journal of Educational Psychology*, 31: 1941.

15) Elizabeth Taylor, "A Study in the Prognosis of Musical Talent," *Journal of Experimental Education*, 10: 1941.

16) E. D. Williams, J. Winter, and J. M. Woods, "Tests of Literary Appreciation," *British Journal of Educational Psychology*, 8: 1938.

17) C. H. Wing, "Tests of Musical Ability and Appreciation," *British Journal of Psychology*, Monograph Supplement, 1948. "A Factorial Study of Music Tests," *British Journal of Psychology*, 31: 1941.

18) Amram Scheinfeld, *You and Heredity*. Chatto & Windus, 1939.

19) L. L. Thurstone, *Multiple-factor Analysis*. Chicago, University of Chicago Press, 1947.

20) G. H. Thomson, *Factorial Analysis of Human Ability*. New York, Houghton-Mifflin, 1939.

THE EFFECT OF INDUCED TENSION DURING TRAINING ON VISUAL FORM RECOGNITION[1, 2]

SYLVIA R. MAYER[3]

Physical Research Laboratories, Boston University

Visual form recognition plays a cental rôle in many educational experiences ranging from the child's early encounter with alphabetical symbols to the adult's identification of biological specimens or aircraft. Visual form learning, which underlies this important skill, is thereby an exceptionally fruitful area for research. Numerous factors in visual learning affect the efficiency of recognition. This study, however, deals only with the effect of induced muscular tension during learning upon subsequent recognition.

Many empirical analyses and speculative accounts appear to support the proposition that induced muscular tension influences learning and that an optimal amount is facilitative. From this proposition two generalizations, which provide the rationale for this study, have been extended to the classroom learning situation. The first states that learning is facilitated by the arousal of a hypothetical condition of "attention" or tension which is both instigated and revealed by observable muscular activity, as for example in the maintenance of erect posture. Support for this view has been drawn from many laboratory studies (e.g., *1*, *3*, *4*, *8*) which have systematically induced tension by procedures such as weight lifting or dynamometer performance. Some of these experiments demonstrate facilitation, but many do not. These equivocal results have recently been partially clarified by a verbal learning study (*2*) which demonstrates that induced tension facilitates performance on a verbal task but has no effect on learning. This finding, however, cannot be generalized to the learning of

---

[1] This work, sponsored by the United States Air Force (Aerial Reconnaissance Laboratory, Wright Air Development Center), was performed by Boston University under Contract No. AF 33-038-15615.

[2] Portions of this paper were presented at the 1955 meetings of the American Psychological Association at San Francisco.

[3] This article reports a part of a dissertation presented to the Graduate School of Boston University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

nonmeaningful visual forms without further qualification. The present study aims to supply the latter.

The second generalization, which is concerned specifically with visual form learning, is an expansion of the first to include this broader notion: the pattern of the tension-inducing activity is important. This notion has been emphasized especially for teaching alphabetical symbols and numerals (*6*, *7*) and biological and cultural forms (*5*, *9*) by means of the traditional activity patterns of drawing and tracing. The assumption appears to be that if S parallels visual activity with isomorphic manual activity, he responds to more form detail than if he engages in visual activity alone. A systematic investigation of this view seems to be needed since the predicted facilitation attributed to these special patterns of activity could as well be the result of a generalized increase in tension level induced by the motor activity *per se*, or the greater amount of training time spent when drawing or tracing are required. The latter possibility is explored in this study.

This experiment is designed to examine the above two generalizations in the context of visual form learning by comparing the effect on recognition of several conditions of training tension. One set of tension conditions is induced by activity which is unrelated in pattern to the visual forms being learned; this is represented by dynamometer performance at several pressure levels. The second set of conditions is induced by activity which is directly related in pattern to the training forms; this is represented by tracing and drawing.

METHOD

Subjects. Twenty-eight men, ranging in age from twenty-two to forty years, served as Ss. Each had 20/20 vision and habitually used his right hand for tasks such as writing and drawing. The group was made up of Air Force officers with various specialties, and technical and research personnel of Boston University Physical Research Laboratories.

Apparatus. The apparatus consisted of four main components: (a) an electronic, dual-controlled tachistoscope for the presentation of visual forms; (b) a Smedley hand dynamometer in circuit with a tachistoscope exposure switch and an auditory signal; (c) graphic materials: 4 x 4 in. white cards for drawing, tachistoscope-aperture covers of plexiglas for tracing, and black grease crayons; and (d)

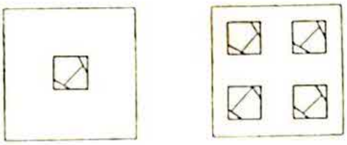

FIGURE 1. An example of a training form (left) and recognition test (right).

TABLE I

Subjects	Presentation Order of Training Conditions						
	1	2	3	4	5	6	7
1, 8, 15, 22	½	0	¼	D	⅛	T	¾
2, 9, 16, 23	0	¼	D	½	T	¾	⅛
3, 10, 17, 24	¼	D	¾	T	½	⅛	0
4, 11, 18, 25	D	¾	T	⅛	¼	0	½
5, 12, 19, 26	¾	T	⅛	¼	0	½	D
6, 13, 20, 27	T	⅛	½	0	¾	D	¼
7, 14, 21, 28	⅛	½	0	¾	D	¼	T

seven ten-item, equated sets of training forms and recognition tests. Each test card contained one form identical to the training form and three similar ones. An example of training forms and tests is shown in Figure 1.

Design. The design used required that each S learn one set of forms in each of the seven conditions: 0, ⅛, ¼, ½ and ¾ maximum dynamometer tension; tracing (T); and drawing (D). The order of presentation of conditions was as in Table I.

Procedure. The seven experimental sessions for each S took place within five days, not more than two a day at a minimum interval of three hours. At each session S performed on one set of forms. In all conditions each of the ten items in a training set was presented for fifteen seconds and spaced at twenty seconds. A ten-minute rest was interpolated between training and testing. The ten items in a recognition test were presented for five seconds each and spaced by ten seconds. S was required to select from four possibilities that form which he had memorized during training. The number of correct recognitions served as the primary response measure.

At the beginning of every experimental session S's maximum dynamometer grip with his left hand was recorded at the end of fifteen seconds of pressure. E then read instructions for the activity of the condition and emphasized the need for careful examination of each form in anticipation of the subsequent recognition test. Three practice forms were presented by the method for the session followed immediately by the recognition test. Correction, additional instructions and practice were provided so that S could achieve perfect performance on the three practice forms.

A three-minute rest was allowed between practice and training. For the ⅛, ¼, ½ and ¾ tension conditions the maximum dynamometer reading of the session was appropriately fractionated and the adjustable tachistoscope switch was positioned on the dynamometer dial at the resulting value. Each form was exposed in the tachistoscope when S exerted the required pressure on the dynamometer handle upon a signal from E. If S overshot or undershot the range of ±three kilograms from the required point, the buzzer sounded until S corrected his grip, but the visual form remained.

In the drawing condition S copied each form onto a card positioned beside the tachistoscope aperture; in the tracing condition he traced each form onto a plexiglas slide placed over the tachistoscope aperture; in the zero-induced tension condition he observed each form, his two arms resting on the tachistoscope stand. Exposures were controlled by E in the latter three conditions. To prevent S from viewing his completed drawing or tracing, a small shutter on the headrest was released at the end of each exposure. During the twenty-second interval E replaced the soiled cards or plexiglas slides with clean ones. The shutter was used in all conditions to maintain experimental uniformity.

RESULTS

Table II presents the mean number of correct responses and standard deviations (SDs) in the seven training conditions. Inspection of this table and Figure 2 shows that (a) increased dynamometer-induced tension is associated with decreased recognition proficiency, (b) drawing is associated with an increase in recognition proficiency over zero-induced tension, and (c) tracing is associated with a decrease in recognition proficiency from zero-induced tension.

The analysis of variance of recognition scores with 0, ⅛, ¼, ½ and ¾ dynamometer induced tension reveals statistically signifi-

TABLE II—RECOGNITION SCORES FOR ALL TRAINING CONDITIONS (N IN EACH CONDITION = 28)

Measure†	Training Condition*						
	0	⅛	¼	½	¾	D	T
Mean	6.14	5.28	5.00	4.75	4.75	7.46	4.78
Standard deviation	1.95	2.30	1.58	1.56	1.36	1.31	2.15

* 0, ⅛, ¼, ½, ¾ maximum dynamometer pressure; drawing (D); tracing (T).

† Recognition scores are based on number of correct responses to 10 test items.

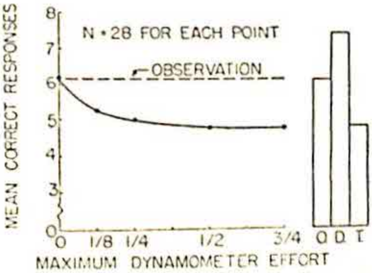

FIGURE 2. The effect of induced tension during training on visual form recognition.

cant differences. The F obtained for tension conditions is 3.97 for four and one hundred and eight df ($p < 0.01$), and the F for individuals is 3.31 for twenty-seven and one hundred and eight df ($p < 0.001$).

Also statistically significant was the analysis of variance for differences between drawing, tracing, and zero-induced tension with an F for conditions of 24.50 for two and fifty-four df ($p < 0.001$), and an F for individual differences of 2.97 for twenty-seven and fifty-four df ($p < 0.001$). Bartlett's chi-square test indicated these analyses to be appropriate.

DISCUSSION

The two generalizations considered in this study are not supported by the results. First, increases in tension level not only failed to facilitate learning but produced marked inhibition. Secondly, the form-related pattern of the tension-inducing activity does not appear to be a major determinant of visual form learning since tracing results in recognition decrement while drawing alone results in improvement in recognition over the zero-induced tension condition.

Evidence from this study suggests that the effect of tension on learning may be determined by the difficulty and coördination requirements of the tension-inducing activity. On the basis of Bourne's findings for verbal learning (*2*), no change in form recognition accuracy was expected as a function of dynamometer induced tension. However, the occurrence of decrement in this study cannot be interpreted as evidence that the effect of such tension is different for visual form learning than for verbal learning since Bourne's experimental procedure differed from this study in an important detail. Bourne required only two-second intervals of pressure at a 1/3-maximum dynamometer tension level as compared to the fifteen seconds of pressure from 1/8 to 3/4 levels in this study. These differences in procedure and results suggest that the effect of dynamometer-induced tension is a function of the difficulty of the procedure: with limited activity, as in Bourne's study, the effect is null; with extensive involvement, as in this study, the effect becomes interfering and hence inhibitory. This speculation is further supported by the occurrence in this study of increased response variation, indicated by dynamometer buzzer signals, which appears to be related to amount of elapsed time within an interval and to the degree of required pressure. Similarly the decrement with tracing may be related to the difficulty of coördination introduced both by the precise manual control necessary to trace the small image, and by the fact that slight head movements result in visual misalignment of image and tracing. In contrast, coördination in drawing is relatively gross and nonprecise.

The facilitation with drawing over the zero-induced tension condition cannot logically be referred to the form-related quality of the tension pattern in view of the decrement with tracing. Instead, a striking feature unique to the drawing condition is suggested as a possible determinant of the facilitation: S necessarily makes successive, discrete visual responses as he glances from the training form to his copy. It may be that this activity promotes an optimal distribution and duration of visual responses as compared to the zero-induced tension condition. This speculation awaits experimental clarification.

SUMMARY

This experiment explores the effect of induced tension during learning on visual form recognition. Experimental conditions differ in terms of kind and amount of tension. In each training condition Ss learn forms which they later recognize from among similar ones.

Recognition proficiency is at a maximum with drawing, at a minimum with tracing, and decreases systematically with increases in dynamometer-induced tension. Evidence is presented that suggests the inhibition phenomena are related to the coördination requirements and difficulty of the tension-inducing activities. The facilitation phenomenon is tentatively referred to the nature of the visual responses accompanying drawing.

REFERENCES

1) A. G. Bills, "The Influence of Muscular Tension on the Efficiency of Mental Work," *American Journal of Psychology*, 38: 227–251, 1927.

2) L. E. Bourne, Jr., "An Evaluation of the Effect of Induced Tension on Performance," *Journal of Experimental Psychology*, 49: 418–422, 1955.

3) F. A. Courts, "Relations between Experimentally Induced Muscular Tension and Memorization," *Journal of Experimental Psychology*, 25: 235–256, 1939.

4) F. A. Courts, "Relations between Muscular Tension and Performance," *Psychological Bulletin*, 39: 347–367, 1942.

5) E. R. Downing, *Teaching Science in the Schools*. Chicago, University of Chicago Press, 1925.

6) A. S. Edwards, *The Psychology of Elementary Education*. Boston, Houghton Mifflin Co., 1925.

7) F. Freeman, *How Children Learn*. Boston, Houghton Mifflin Co., 1917.

8) G. L. Freeman, "The Optimal Muscular Tensions for Various Performances," *American Journal of Psychology*, 51: 146–150, 1938.

9) J. J. Gibson, editor, *Motion Picture Testing and Research*. AAF Aviation Psychology Program Research Report No. 7, 1947.

CORRELATION ANALYSIS OF WISC AND ACHIEVEMENT TESTS

JAMES B. STROUD and PAUL BLOMMERS

State University of Iowa

MARGARET LAUBER

State of Iowa Department of Public Insruction

The *Wechsler Intelligence Scale for Children* (WISC) has within its short history established itself as a popular test. It is now being used extensively in school psychological work. Yet insofar as the authors can ascertain there has been no large-scale investigation of the performance of this test in school situations.

The primary purpose of this investigation was to determine the effectiveness with which all or various combinations of the WISC subtests could be used to predict performance on the Reading Comprehension, Arithmetic, and Spelling tests of the Iowa Tests of Basic Skills (ITBS) battery. The sample consisted of seven hundred and seventy-five pupils in Grades three to six. Stanford-Binet (S-B) I.Q.'s were also available for six hundred and twenty-one of these pupils, making possible comparisons between WISC and S-B as predictors of performance on the three achievement tests for a large sample of elementary school pupils.

The pupils in this sample were drawn from a twenty-county area in Iowa. The numbers of pupils from cities, towns, consolidated and rural schools were roughly proportional to the distribution of pupils in these categories. The WISC tests were administered by Lauber in connection with her work as a regional school psychologist for the twenty-county area. The ITBS battery was administered as a part of the University of Iowa's state-wide testing program. Some of the S-B tests were administered by Lauber, but a majority were administered by local school psychologists and supervisors. The tests were administered over a three-year period, beginning with the school year 1952–53. The ITBS battery was administered in all participating schools in January of each year. The WISC tests were administered to the pupils in this sample within three months (before or after) of the date on which they took the ITBS. This time relationship was not observed in the case of the S-B. Such S-B scores as were available were utilized, regardless of the time at which the test was administered.[1]

The pupils in this sample were all "referred" by teachers and/or school officials for psychological interviews and examinations. All of them were in, or were thought to be in, some kind of school difficulty. The mean Full Scale WISC I.Q. of the girls in this sample was 87.3; that of the boys, 89.7; and that of both sexes, 89.0. The within-grades S.D. of the I.Q.'s was 14.0 for girls and 14.8 for boys, and the within-grades within-sexes S.D. was 14.51. The mean I.Q. of the sample is considerably below average, although the S.D. approaches the value for the general population. The pupils in the sample tend to be somewhat overage for their respective grades. The mean chronological ages by grades at the time WISC was administered were as follows: Grade three, CA 9:1; Grade four, CA 10:2; Grade five, CA 11:7; Grade six, CA 12:7. Also the pupils in the sample were, on the average, at least a year retarded in the three achievement areas studied. Perhaps it is reasonable to suppose that this sample is representative of the population to which we would most commonly wish to generalize in school practice. Pupils presenting some kind of school problem make up the bulk of those receiving individual psychological examinations.

As has just been indicated, all subjects studied were referrals. The problem categories and frequency of classification, as taken from the regional school psychologist's files, were as follows for the two sexes combined: Mental Retardation, sixty-three; Educational Retardation (not mentally retarded, but requiring remedial instruction, especially in reading, and teacher's assistance in structuring learning activities), two hundred and sixty-eight; Personality Deviations, apparently of long standing, twenty-five; Transient Personality Deviations (although pupil may react rather violently to frustrations, he responds to improved environment), one hundred and sixty-one; Emotional Problems, evincing serious disturbance, one hundred and eighty-eight; miscellaneous, some showing at least superficial symptoms of brain damage, seventy.

---

[1] For a random subsample of one hundred and fifty-six the mean time separation for the two administrations was two years and 10.4 months. In every case in this subsample S-B was administered first.

PROCEDURE

(1) Within-grades zero order correlations were computed, by sex and within-sexes, between WISC Verbal, Nonverbal, and Full Scale I.Q.'s and each of the achievement tests, and between S-B I.Q.'s and the achievement tests.

(2) Within-grades intercorrelations, by sex and within-sexes, were computed involving S-B I.Q.'s and WISC Verbal I.Q.'s, Nonverbal I.Q.'s, and Full Scale I.Q.'s.

(3) Within-grades zero order intercorrelations were computed among WISC subtest raw scores and the achievement tests.

(4) *Beta* weights for a within-grades regression analysis involving Reading, Arithmetic and Spelling as dependent variables and WISC subtest raw scores as independent variables were computed, and tested for significance.[2] The nonsignificant weights were eliminated in turn, beginning with the weight for which the t ratio was least, and regression equations determined. Multiple correlations between the achievement tests and various combinations of the WISC subtests were computed.

(5) In order to test the predictions of the various regression equations on a sample other than that from which they were derived, a cross-validation study was carried out on an independent sample of one hundred and ninety-nine pupils drawn in the same way as the original sample.

RESULTS

(1) The within-grades correlations, by sex and within-sexes, between WISC Verbal I.Q.'s, Nonverbal I.Q.'s, Full Scale I.Q.'s, S-B I.Q.'s and each of the three achievement tests are presented in Table I.

It is noteworthy that the r between WISC Nonverbal I.Q.'s and reading was somewhat higher than that between Verbal I.Q.'s and reading. The difference is statistically significant. With arithmetic and spelling the Verbal I.Q.'s yielded higher r's than the Nonverbal I.Q.'s. The difference is significant only in the case of arithmetic. WISC Full Scale I.Q.'s correlated with each of the achievement test scores to a higher extent than did S-B and ITBS. Although the differences are significant, they should be interpreted

---

[2] A five per cent significance level was used throughout this investigation.

TABLE I—WITHIN GRADES CORRELATIONS BETWEEN ITBS SCORES AND VARIOUS MEASURES OF INTELLIGENCE

Test Variables	Females: N = 205			Males: N = 416			Within Sexes: N = 621		
	Reading	Arithmetic	Spelling	Reading	Arithmetic	Spelling	Reading	Arithmetic	Spelling
W: V I.Q.	0.59	0.65	0.55	0.57	0.70	0.66*	0.58	0.67	0.62
W: Nv I.Q.	0.67	0.55	0.56	0.62	0.50	0.62	0.63	0.52	0.60
W: V + Nv I.Q.	0.69	0.65	0.60	0.65	0.66	0.71*	0.66	0.66	0.67
S-B I.Q.	0.59	0.58	0.51	0.59	0.63	0.66*	0.59	0.61	0.61

* Difference between corresponding r's for sexes statistically significant. All other differences between sexes were nonsignificant.

in the light of the time interval between administrations of S-B and ITBS. As noted earlier, the maximum time interval between the administration of WISC and ITBS was only three months. The approximate mean time interval between administration of S-B and WISC was almost three years. This average is also indicative of the time interval between administrations of S-B and ITBS.

(2) Within-grades intercorrelations, by sex and within-sexes, between the various I.Q.'s are given in Table II. The magnitude of the correlation (r = 0.94) between WISC Full Scale I.Q.'s and S-B I.Q.'s was a little greater than one might have anticipated in view of the time interval between the administration of the two tests.

(3) Within-grades intercorrelations among WISC subtest raw scores and ITBS scores are shown in Table III. It may be seen

TABLE II—WITHIN GRADES CORRELATIONS BETWEEN VARIOUS MEASURES OF INTELLIGENCE

Test Variables	Females: N = 205			Males: N = 416			Within Sexes: N = 621		
	W: Nv I.Q.	W: FS I.Q.	S-B I.Q.	W: Nv I.Q.	W: FS I.Q.	S-B I.Q.	W: Nv I.Q.	W: FS I.Q.	S-B I.Q.
W: V I.Q.	0.73	0.92	0.87	0.64	0.91	0.87	0.67	0.91	0.87
W: Nv I.Q.		0.93	0.85		0.90	0.83		0.91	0.83
W: FS I.Q.			0.93			0.94			0.94

Bureau Edn. Research COLLEGE

TABLE III—WITHIN GRADE INTERCORRELATIONS: WISC AND ITBS
(N = 775)

Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. W: Information		0.48	0.49	0.47	0.60	0.13	0.25	0.37	0.31	0.29	0.21	0.40	0.44	0.40
2. W: Comprehension			0.40	0.50	0.53	0.15	0.34	0.41	0.39	0.36	0.14	0.43	0.44	0.41
3. W: Arithmetic				0.39	0.46	0.20	0.28	0.30	0.36	0.34	0.24	0.46	0.70	0.50
4. W: Similarities					0.49	0.15	0.33	0.31	0.35	0.29	0.27	0.42	0.43	0.41
5. W: Vocabulary						0.08	0.27	0.39	0.30	0.11	0.20	0.41	0.45	0.41
6. W: Digit Span							0.12	0.15	0.14	0.13	0.13	0.20	0.26	0.23
7. W: Picture Comp.								0.34	0.36	0.28	0.16	0.29	0.26	0.28
8. W: Picture Arrangement									0.36	0.44	0.24	0.43	0.36	0.39
9. W: Block Design										0.50	0.25	0.57	0.45	0.51
10. W: Object Assembly											0.26	0.59	0.41	0.51
11. W: Coding												0.31	0.29	0.31
12. ITBS: Reading Comp.													0.73	0.84
13. ITBS: Arithmetic														0.77
14. ITBS: Spelling														

that Digit Span yielded comparatively low r's not only with the three achievement tests but also with the other subtests. This test did, however, in subsequent analyses yield significant *beta* weights[3]. The r's between both Picture Completion and Coding and the achievement tests were also relatively low. Coding, like Digit Span, yielded significant *beta* weights, but the weights for Picture Completion were nonsignificant. The r's between Block Design and reading and between Object Assembly and reading are exceptionally high. These two r's are significantly larger than the r's between reading and all other WISC subtests. The r's between spelling and WISC Block Design, Object Assembly and arithmetic were significantly higher than those between spelling and all other subtests. The r's between arithmetic and WISC Digit Span, Picture Completion, and Coding were significantly lower than those between arithmetic and all other WISC subtests except Picture Arrangement; as would be expected, the r between arithmetic and WISC Arithmetic was significantly higher than any of the r's between this variable and all other WISC subtests.

In general the intercorrelations among the WISC subtests for this sample were lower than those reported by Wechsler. However, the relative magnitudes (i.e., rank order) of the coefficients were practically the same as obtained by Wechsler.

---

[3] The results of the regression analyses are presented in (4).

(4) Multiple regression analyses based on the within-grades intercorrelations reported in Table III were performed with ITBS Reading, Arithmetic, and Spelling, as dependent variables, and all WISC subtest raw scores, as independent variables. *Beta* weights were tested for statistical significance and the non-significant variables successively eliminated. The six tests which carried significant weights in predicting achievement in reading and spelling were WISC Arithmetic, Digit Span, Vocabulary, Block Design, Object Assembly, and Coding. In the case of arithmetic achievement these six tests were joined by WISC Comprehension. Among these subtests Comprehension (in the case of the arithmetic criterion), Coding, and Digit Span were actually of negligible *practical* significance even though statistically significant. The *beta*-weights and multiple correlation coefficients for various combinations of WISC tests are shown in Table IV.

It is of particular interest that for all practical purposes the same set of WISC tests (Arithmetic, Vocabulary, Block Design, and Object Assembly) are the most effective predictors of each achievement test investigated. Moreover, the R's obtained between the least squares composites of these four subtests (raw scores) and each of the criterion tests were practically as high as those based on all eleven tests (raw scores) and the criterion tests. Also the r's between the unweighted sums of the raw scores of these

Table IV—Multiple R's and *beta*-Weights for within-Grades Regression Analyses (N = 775)

Multiple R's	Reading			Arithmetic			Spelling		
	.74	.74*	.73*	.76	.76*	.75*	.70	.60*	.68*
Inf. β	−0.01			−0.01			0.01		
Comp. β	0.02			0.06	0.07		0.01		
Arith. β	0.11	0.12	0.13	0.50	0.50	0.53	0.20	0.20	0.23
Sim. β	0.07			0.06			0.05		
Vocab. β	0.19	0.22	0.22	0.09	0.11	0.14	0.17	0.20	0.21
D.S. β	0.06	0.06		0.10	0.10		0.10	0.10	
P. Com. β	−0.04			−0.04			−0.02		
P. Ar. β	0.04			0.03			0.03		
D.B. β	0.25	0.25	0.26	0.13	0.12	0.14	0.20	0.21	0.22
O.A. β	0.36	0.38	0.40	0.10	0.11	0.14	0.26	0.28	0.30
Cod. β	0.06	0.07		0.06	0.06		0.08	0.09	

* All weights statistically significant.

TABLE V—WITHIN GRADES CORRELATIONS OF VARIOUS TYPES BETWEEN ITBS TESTS AND SELECTED COMBINATIONS OF WISC TESTS

WISC Tests	N = 775 Least Sqs. Multiple (Raw Scores)	N = 775 Unwtd. Sums (Raw Scores)	N = 621 Unwtd. Sums* (Scale Scores)	N = 621 WISC I.Q.'s
Reading				
Eleven W: tests (6V + 5Nv)	0.742	0.69	0.65	0.66
Verbal		0.52	0.59	0.58
Nonverbal		0.66	0.63	0.63
3, 5, 6, 9, 10, 11†	0.74	0.68		
3, 5, 9, 10	0.73	0.69		
Arithmetic				
Eleven W: tests (6V + 5Nv)	0.76	0.63	0.66	0.66
Verbal		0.61	0.68	0.67
Nonverbal		0.53	0.51	0.52
2, 3, 5, 6, 9, 10, 11	0.76	0.63		
3, 5, 9, 10	0.75	0.61		
Spelling				
Eleven W: tests (6V + 5Nv)	0.70	0.65	0.68	0.67
Verbal		0.55	0.63	0.63
Nonverbal		0.60	0.60	0.60
3, 5, 6, 9, 10, 11	0.69	0.65		
3, 5, 9, 10	0.68	0.64		

* Verbal prorated at 5/6.

† 2 = Comprehension
3 = Arithmetic
5 = Vocabulary
6 = Digit Span
9 = Block Design
10 = Object Assembly
11 = Coding

TABLE VI—COMPARISONS BETWEEN WITHIN-GRADES CROSS-VALIDATION R's (N = 199) AND R's FOR ORIGINAL SAMPLE (N = 775)

Criterion	WISC Tests	R's Cross Validation	R's Original Sample
ITBS: Reading	3, 5, 6, 9, 10, 11	0.69	0.74
	3, 5, 9, 10	0.69	0.72
ITBS: Arithmetic	2, 3, 5, 6, 9, 10, 11	0.68	0.76
	3, 5, 9, 10	0.69	0.74
ITBS: Spelling	3, 5, 6, 9, 10, 11	0.65	0.69
	3, 5, 9, 10	0.66	0.68

four subtests and the criterion tests are practically as high as those between the unweighted raw score sums for all eleven tests and the criterion tests (see Table V). As a further step, I.Q.'s were determined for these four tests from the WISC I.Q. table upon the basis of 10/4 times the sum of their scale scores. The r between the I.Q.'s thus derived from this abridged battery and the I.Q.'s derived from the complete battery (eleven tests) was 0.85. The r between Verbal I.Q.'s (six tests prorated) and I.Q.'s based upon 5/2 times the sum of the scale scores on WISC Arithmetic and Vocabulary was 0.83; that between Nonverbal I.Q.'s (five tests) and I.Q.'s based upon 5/2 times the sum of the scale scores on Block Design and Object Assembly was 0.82.

The r's between (1) WISC raw score sums, (2) WISC scale score sums, and (3) WISC I.Q.'s and the three achievement tests are given in the last three columns of Table V. To facilitate comparison, the R's shown in Table IV are repeated in Column 1 of this table. As was noted above, it may be observed that the r's between the sums of unweighted raw scores of WISC Tests 3, 5, 9, and 10 and the three achievement tests are relatively high. This suggests that the school practitioner might do about as well to administer WISC Tests 3, 5, 9, and 10, multiply the scale score sums by 10/4 and consult the table for conversion of full scale scores into WISC I.Q.'s, as to proceed in the conventional way. This, of course, is on the assumption that the scores made on these four subtests are unaffected by taking the other tests in the battery. The assumption may not be warranted. To obtain definitive evidence one should have to administer these four subtests as a four-test battery.

(5) *Cross-validation.* The data in the preceding section tell us how the regression equations perform for the sample from which they were derived. More important is the way they perform, once derived, for an independent sample. To ascertain this, an independent sample of one hundred and ninety-nine pupils was drawn from the same population as the original sample. Within-grades correlations were computed between various combinations of WISC tests[4] and the three criterion measures. The results are presented in Table VI. For convenience in comparison, the last column of Table VI shows the multiple R's between various combinations of WISC subtest unweighted raw scores and the three criterion tests for the original sample. Over all, only a small shrinkage resulted. The average difference between the correlations for the original sample and the cross-validation sample is about 0.04.

Comments. In this investigation WISC gave evidence that its popularity is well-deserved. The results justify the wide use of this test in school situations. Some of the subtests, however, seem to contribute very little to the prediction of school achievement as measured. It would appear, therefore, that serious attention should be given to the possibility of developing an abbreviated battery. Our results suggest the use of Arithmetic, Vocabulary, Block Design, and Object Assembly. It is recognized that such an abbreviated scale would have limited use. There is no intent to imply that it should be substituted for the full test in all situations.

The WISC in keeping with current practice in test design provides both verbal and nonverbal scores and subtest scores in each category, so as to permit differential prediction. Insofar as our analyses of data go, there does not appear to be a great deal of support for this practice with respect to the WISC tests. The total score, as represented by Binet's concept of a "general average" is still far and away the most revealing score, so far as prediction of academic achievement is concerned.

---

[4] The least squares weights based on the original sample were used.

A COMPARISON OF INSTRUCTION BY KINESCOPE, CORRESPONDENCE STUDY, AND CUSTOMARY CLASSROOM PROCEDURES

THOMAS S. PARSONS

University of Michigan

Growing interest in the integration of television into universities, on- and off-campus instructional programs has begun to pose many questions concerning the efficacy of "canned" kinescope presentations in relation to more conventional methods. Husband (*1*), in a comparison of live TV, kinescope, and conventional classroom presentations of similar course material, has produced results indicating superior achievement for the kinescope presentation. However, his report does not reveal the statistical significance of this superiority, the relative size of his classroom, kinescope, or TV-viewing groups, nor statistical evidence of the original comparability of his experimental groups. Evans, Roney, and McAdams (*2*), in a comparative study of achievement scores for groups taught by conventional lecture, TV-plus-correspondence, and TV-plus-classroom discussion methods, found no statistically significant differences among any of these groups.

The present paper compares kinescope-correspondence study groups without an instructor, conventional classroom discussion experience with an instructor, and independent correspondence study for their relative effectiveness as methods of providing university-level instruction in identical subject matter (psychology of child development). Course effectiveness is measured before, during, immediately after, and approximately four months following the course in terms of academic achievement, certain group processes, and students' attitudes toward the experience.

SUBJECTS

Forty university upperclassmen served as subjects. Twenty students enrolled in the experimenter's developmental psychology course were randomly assigned to one of three groups, each of which received instruction according to one of the experimental treatments described above. The kinescope group was subdivided into two small viewing groups. Twenty students concurrently enrolled in the experimenter's educational sociology course were given no consistent instruction in the experimental subject matter and were used as controls.

PROCEDURES

Common to all three experimental treatments (but not to the control group) was the course textbook (*3*), a syllabus-workbook including fourteen written assignments to be completed and graded during the course, and several auxiliary manuals covering specific topics such as sociometry. Identical subject matter was thus presented in the same sequence in each group. The classroom group, however, met with the experimenter at regular class sessions twice a week, whereas the kinescope groups each met approximately once a week in a drawing room to view filmed TV lectures on the current course topics. Members of the correspondence group worked entirely independently, viewed no filmed presentations, and—like the kinescope group—maintained only postal contact with the instructor through the fourteen written assignments and the midsemester examination. These were regularly corrected and returned with appropriate comments. The experimenter served as instructor both in the classroom and through correspondence—rôles in which he had had substantial experience prior to this study.

The control group was compared with the experimental groups initially, midway, and at the end of the study for relative achievement of the experimental subject matter. The three experimental groups were compared laterally for between-groups differences—and longitudinally for within-groups differences whenever appropriate—in seven dependent variables. These, as suggested above, included achievement, group cohesiveness, congruence of course-associated opinions, the degree of perceived intragroup structure or interpersonal dependability, preferences among the three experimental conditions of instruction, composite grades for the course, and subjects' evaluations of this course's personal value as compared with others taken concurrently. Data were collected prior to active instruction, halfway through the course, and immediately following the course by use of standard multiple-choice achievement tests, a modified Libo-Schachter cohesiveness questionnaire, a sociometric test ("With whom would you most (least) like to discuss topics covered in the course?"), and *ad hoc* questionnaires and rating scales. In addition, a "retention" achievement test, sociometry, and questionnaires were administered about four months after the course. Analyses incorporated chi-square, t, and the analysis of variance as measures of the significance of observed differences, and C and Φ—where applicable—as crude measures of relationships among the independent and dependent variables.

RESULTS

Achievement. (See Figure 1.) No significant differences in achievement appeared among any groups at the outset, nor among the three experimental groups' scores on the midsemester. At the end of the semester, however, the difference in achievement between the experimental groups and the control group was significant at P = 0.001; whereas there remained no significant differences among experimental groups. Similarly, scores based on an evaluation of all written work done throughout the course showed no significant differences among the experimental groups.

Four months after the course, however, a difference just barely significant at P = 0.05 was found which favored the correspondence group over the classroom group, although no significant differences appeared at this time between the classroom group and the kinescope group or between the kinescope group and the correspondence group. A comparative analysis of the *net gains* in achievement

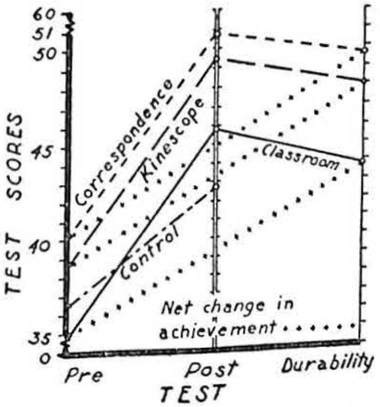

FIGURE 1. Mean achievement of the three experimental groups on the pre-experimental, post-experimental, and retention tests.

retained by the three experimental groups four months after the course ("retention" test scores minus precourse test scores) yielded an F ratio which was definitely not statistically significant; and a comparison of the three groups for differences in the amount of post-course change in tested achievement (post-course test scores minus retention test scores) produced an F ratio which was similarly far from statistical significance.

Group cohesiveness. Members of the correspondence group were never administered the cohesiveness questionnaire since its contents would have been meaningless to them; their actual cohesiveness as a group was assumed to be nil. Cohesiveness scores were obtained from the classroom and kinescope groups, however, and means for these two groups were not significantly different at any point during the semester. When the classroom group's mean was adjusted by eliminating one highly deviant score on the post-course measurement (produced by a psychiatrist's daughter) it rose considerably—above the mean for the kinescope group. Yet, even after this adjustment the difference between the two groups at this point remained far from significant.

Congruence of course-associated opinions. (See Figure 2.) On the assumptions that friendly interaction among individuals tends to cause their opinions about the central topics of interaction to converge—and that this convergence of opinions is as likely to occur for invalid or wrong opinions as for valid opinions—measurements of the congruence of course-associated opinions were derived

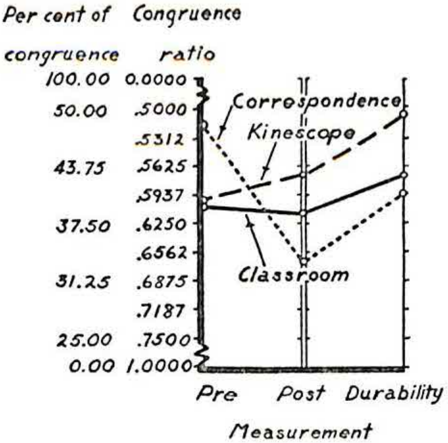

FIGURE 2. Convergence of course-associated opinions for the three experimental groups on the pre-course, post-course, and durability examinations.

for the three experimental groups from detailed item analyses of their initial, final, and "retention" achievement tests. These measures were expressed as a "congruence ratio":

$$\frac{\text{Sum of multiple-choice categories in which incorrect answers were given}}{\text{Sum of incorrect answers}} = \frac{c}{i}$$

and the ratios for the three groups were examined for longitudinal differences within groups and for differences between groups.

Although a statistically significant ($P < 0.05$) decrease in congruence was indicated for the correspondence group during the course, no significant longitudinal changes were found for the other two (presumably high interaction) groups either during or after the course, or for the correspondence group after the course. Similarly, no significant differences in opinion congruence were found among any of the groups at any point when compared laterally, even though the small kinescope groups were consistently highest and the interactionless correspondence group was consistently lowest in opinion congruence both during and after the course.

Group structure. (See Figure 3.) On the assumption that group members' willingness to choose others as performing or not performing functions central to the group's purpose is a correlate of their perceptions of the group's structure, measures of group structure or the dependability of rôle expectations were derived from composite (positive plus negative) sociometric subject scores for the classroom and kinescope groups. These showed a significant ($P = 0.05$) increase during the semester for the classroom treatment and a decrease which was not quite significant at $P = 0.05$

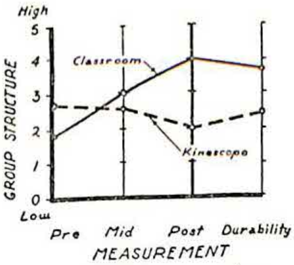

FIGURE 3. Mean group structure scores (sum of "choice" and "rejection" sociometric subject scores) for the classroom and kinescope groups on pre-course, mid-course, and durability measurements.

for the kinescope treatment. Similarly, lateral or between-groups comparisons of the classroom and kinescope groups on this variable indicated no significant differences at the beginning or middle of the semester, but a difference significant at beyond P = 0.01 by the semester's end. The delayed comparison of these two groups for durability of effect—or retention—showed that the kinescope group's sociometric choices had increased and the classroom group's had decreased, thus bringing them closer together than at the end of the course in their potentials for interpersonal dependability. Nevertheless, even at this point the difference between the two groups in this variable remained significant at far beyond P = 0.05.

A source of error in this comparison of group structure was introduced by the fact that for the kinescope group n = 7, while for the classroom group n = 6, thus consistently giving the kinescope group a greater chance to make sociometric choices than the classroom group. However, any adjustment for this disparity would tend to lower the kinescope group's or raise the classroom group's indicated structure, thus uniformly increasing the between-groups differences described above.

Methods preferences. (See Figures 4a, b, and c.) Students' prefer-

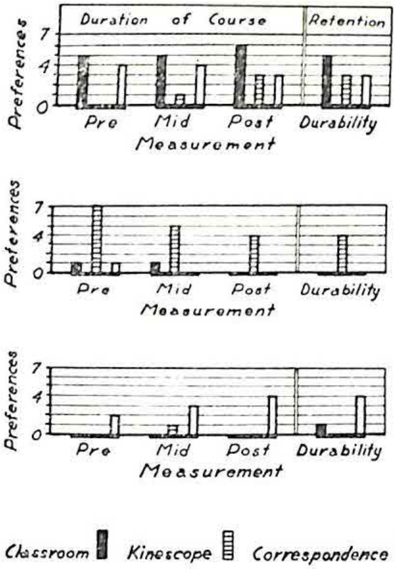

FIGURE 4. Methods preferences for the three experimental groups on the pre-course, post-course, and durability measurements. a (upper), choices for classroom (accustomed experience); b (middle), choices for kinescope (partly unaccustomed experience); c (lower), choices for correspondence (totally unaccustomed experience).

ences among the three experimental instructional treatments—as measured both prior to and immediately following the course—bore a stable, highly significant positive relationship to the treatments to which they had actually been arbitrarily assigned. (At both times $C = +0.80$, $\Phi = +0.86$; $P = 0.001$.) Almost all individual statements of preference which deviated from this tendency were accounted for by members of the kinescope and correspondence groups who preferred the classroom experience, although this countertendency progressively *decreased* throughout the course for the correspondence group and tended to *increase* during the course for the kinescope group.

Four months after the course this high relationship remained stable and substantially unaltered ($C = +0.78$, $\Phi = +0.82$; $P = 0.01$), although some reduction in reliability resulted from the use of combined measures for the analyses of pre- and post-course preferences[1] while simple measures were used in the analysis of retained preferences. Once again, almost all countertendency to the obtained relationship was contributed by an equal number of students in the correspondence and kinescope groups who stated preferences for the classroom treatment, thus stabilizing the performance of these two groups on the immediate post-course measurement and ending their respective trends toward and away from their own experimental treatments.

Course evaluations. The students' immediate post-experimental evaluations of this course in comparison to others taken concurrently revealed that on a three-point scale from "more valuable" (1) to "less valuable" (3) the mean ratings were 1.0 for the classroom group, 1.7 for the kinescope group, and 1.4 for the correspondence group.

Analysis of the relationship between these evaluative ratings and the subjects' experimental treatments—arranged in order of most (classroom) to least (correspondence) similarity to their accustomed educational backgrounds—yielded a small positive relationship ($C = +0.60$, $\Phi = +0.56$) which was not significant at $P = 0.05$. Four months later these ratings had not significantly altered (classroom group $M = 1.2$; kinescope $M = 1.7$; correspondence $M = 1.7$), nor had the relationship between them and the students' instructional treatments changed remarkably ($C = +0.44$, $\Phi = +0.38$; $P > 0.50$). However, a tendency appeared for students in all but the kinescope group (which was rated lowest immediately after the course) to lower their ratings slightly in the direction of "about average value."

---

[1] On the assumption that the mid-course measurements would serve as the best estimate of the mid-point of change during the course, the pre- and post-course analyses of methods preferences were based on n^s of thirty rather than of twenty. These were obtained by using pre-course measures plus one-half of the mid-course measures for the pre-course analysis, and by using post-course measures plus one-half of the mid-course measures for the post-course analysis. Additional tests were run with uncombined data to determine the amount of spurious correlation introduced by this multiple use of individuals' preferences, and no substantial amount was found.

CONCLUSIONS

Throughout the following discussion it should be borne in mind that the findings presented above were based upon experimental manipulations which involved never more than forty subjects. Thus, even though subjects were randomly distributed among experimental groups and experimentally produced differences were usually tested by the potent F ratio, any interpretation of these findings may not inspire the degree of confidence which would be possible if a larger N had been subjected to similar procedures.

Therefore, *to the extent* that we can consider these findings a reliable representation of adult students' reactions to the three experimental treatments it may be concluded that:

(a) None of the three instructional methods is significantly more effective than the others in producing terminal achievement. All of the methods, as might be expected, are better in this respect than exposing the student to no systematic instruction at all. In addition, a literal interpretation of this analysis would indicate that the post-instructional retention of pure factual achievement tends very slightly to favor the correspondence method—and perhaps the kinescope method—over the classroom treatment. (See Figure 1.) Thus, *if one is willing to consider pure tested achievement an adequate index of the desired outcomes of instruction*, it tentatively appears that kinescope (or television) instruction is as effective as class discussion methods for the teaching of abstract or highly verbal subject matter. Furthermore—and this may be of interest to those contemplating the expense of educational TV transmission—it seems that independent correspondence study is at least as effective in this respect as kinescope instruction, and probably more so than the classroom. Some may feel that this finding suggests an expedient solution to the problem of overcrowded classrooms and understaffed teaching faculties.

The reason for this superiority of the correspondence method may not be hard to find. Certainly the correspondence group in particular had no opportunity to be distracted from the factual content of the text (on which the examination was largely based) by the "interesting implications," "personal applications," and group interaction phenomena with which the class members were bombarded. On the other hand, a close examination of the data indicates that this significant difference in retained achievement may be due in part to a progressive spreading of group achievement means which were already separated (though not quite significantly) at the outset. The fact that there were almost no differences among the groups' *net gains* in achievement during the course and among their post-course losses in achievement would tend to support this latter interpretation.

Finally, one might argue that tested achievement, alone, is a meager measure of the total desired outcomes of instruction. Such an argument would lead to a comparative examination of personal satisfaction, of the amount of opinion change, or of certain group interaction variables before a final conclusion could be drawn concerning relative instructional effectiveness.

(b) The amount of group structure, cohesiveness, acceptance of others' opinions, and certain other presumed outcomes of group interaction (see Figures 2 and 3) show no statistically reliable differences between the classroom and kinescope treatments during the course except for what appears to be a very reliable (and conservative) indication that the classroom-with-instructor method tends to produce group structure or dependable rôle expectations among members at a faster rate and to a higher degree than does the instructorless kinescope method (Figure 3). And this difference is quite durable; for even though a convergent or regressive effect is produced by the lapse of time following active instruction, the classroom group still retains significantly more rôle interdependability than the kinescope groups after approximately four months of separation.

It could be argued that the classroom group's perceived structure is narrowly related to the course's formal content, and that if the kinescope group met as many hours as the classroom group, or if this variable were measured by a sociometric criterion broader than "With whom would you most (least) like to discuss topics covered in the course?" the obtained superiority of the classroom group might be nullified or reversed. Such a finding, in fact, would be much easier to reconcile theoretically with the kinescope group's slight superiority in both cohesiveness and opinion convergence. It would rectify the appearance given here of a negative relationship between these two variables and the structuredness of group members' rôle expectations. Nevertheless, even though the classroom instructor may deliberately establish channels for intercommunication on matters predominantly concerning the course's formal content—whereas members of the instructorless group might spend more of their time in attempting to establish a more universally applicable set of interpersonal rôle expectancies—it would seem that the *academic* purposes of any instructional group which perceives itself as such must remain the hub of both intragroup communication and of the rôle expectancies which become central to the group's processes. In addition, the theoretical objection raised above is considerably weakened when one recalls, first, that the omission of a single highly deviant cohesiveness score raises the classroom group substantially above the kinescope group on this variable, and second, that between-groups differences in both cohesiveness and opinion convergence were not statistically significant either before, during, or after the course. To the objection that the meeting hours of the kinescope group were considerably fewer than those of the classroom group it may be answered that this disparity of "contact hours" has been a common characteristic of the kinescoped or televised reproduction of a college course, and that this comparison is still, therefore, a valid one. Thus the classroom treatment's superiority as a method for producing durable interpersonal rôle expectancies or perceived structure appears to stand up under both logical and statistical analyses.

A final note concerning the rough validity of these group interaction measures is that they appear to be supported by the interactionless correspondence group's significant decrease in opinion convergence throughout the course, and its position as the lowest of the three groups in this variable by the end of active instruction (Figure 2). Following the course, when contact was also lost among members of the formerly "high interaction" classroom and kinescope groups, all three groups' measured opinion convergence rose—probably as a result of practice effect.[2] However, the correspondence group remained consistently below the others in this variable even though its post-course rate of gain was greater.

(c) Students' preferences among instructional methods appear to be mightily conditioned, first, by what they are immediately involved in, and secondly, by what they have become accustomed to in the past (Figures 4a, b, and c). In all cases the factor of immediate involvement is the most powerful determinant of methods preferences. Thus students in all three experimental groups stated a preference for the treatment to which they had been arbitrarily assigned far more than for any other treatment or treatments. However, where the immediate experience differs in relatively small degree from the accustomed instructional pattern (as in the kinescope group) the initial appeal of novelty soon diminishes for some students and the factor of custom ultimately becomes more important in determining preferences. Conversely, to the extent that the immediate experience is more essentially different from customary patterns (as in the correspondence treatment) this experience appears to become incomparable with these patterns—to be evaluated according to new standards; and the factor of immediacy or present involvement becomes relatively more important. Where the factors of immediacy and custom are working in the same direction (as in the classroom treatment) there emerges an apparently unanimous preference for the immediate experience as opposed to any alternatives.

The stability of this relationship appears to be great. For once these two factors—custom and immediacy—have interacted and established a balance, this balance appears to resist change long after the experiences which produced it have ceased to exist. It may be that the relative appeal of these two factors is determined by two personality syndromes which remain stable and separate within the student population. Here, again, may be a point of interest to those who debate the relative merits of TV or correspondence study as substitutes for overcrowded and understaffed classrooms.

---

[2] The same sixty-item, four-alternative, multiple choice achievement test was used in the pre-course, post-course, and durability measurements. This test as a whole probably became increasingly familiar to all three groups; and in addition, as knowledge of course content grew, more and more alternatives per item were probably immediately ruled out as possible correct answers. Thus, during the four months' retention period—when all groups again became equal in the interaction variables—the greatest gain in opinion convergence was made by the correspondence group, the next greatest by the kinescope group, and the least by the classroom group. This is exactly the order in which the three groups fell in tested achievement both at the beginning and end of the course and after the retention period.

Students' end-of-course evaluations of its personal value to them also appeared to be conditioned (though less powerfully and far less reliably) by the perceived comparability of their groups' instructional treatments to customary patterns of instruction. Thus, in their immediate post-experimental ratings the "traditional" classroom group unanimously rated the course more valuable than others taken concurrently, while the "novel" correspondence group's mean rating was about halfway between greater and average value, and the "intermediate" kinescope group's mean rating was little better than "of average value." The effect of time passage following an instructional experience appears to be a leveling, or a regression toward mediocrity in the student's memory. This is indicated by the durability measures for both the classroom and correspondence groups, although the previously low kinescope group made no post-course change at all.

A final inference drawn from these findings might suggest that—contrary to what seems to be popular practice—the greatest chances for the success of educational TV (or any new instructional medium) may lie in the unabashed exploitation of the medium's intrinsic advantages—precisely, the extension of detailed visibility—rather than in the cautious aping of tried and proven classroom techniques.

SUMMARY

Kinescope-correspondence instruction apart from an instructor, customary classroom experience, and independent correspondence study were compared for their relative effectiveness as methods of providing university-level instruction in identical subject matter. Subjects were randomly assigned among the three experimental treatments and a control group was used. All groups were compared in achievement, and comparisons were made among the three experimental groups, only, on congruence of course-related opinions, preferences among the methods, and subjects' ratings of the course's value to them. Comparisons were made between the kinescope group and classroom group, only, on cohesiveness and group

structure. Measurements of these variables were made prior to instruction, midway in the course, and after a four months' time lapse.

Throughout the course it was found that although no significant differences appeared among the experimental groups' achievement, cohesiveness, and ratings of the course's personal value, all three experimental groups achieved very significantly more and faster than the control group. In addition, comparisons of the amount of congruence of content-related opinions indicated a significant decrease in congruence for the correspondence group, but no significant changes for the other two treatments. Measures of the degree of group structure showed a significant increase during the semester for the classroom treatment and a not quite significant decrease for the kinescope group; thus, comparisons of these two groups at the end of the course revealed a highly significant difference between them. The subjects' preferences among experimental treatments bore a highly significant positive relationship to the treatments to which they had been assigned; however, almost all preferences which deviated from this tendency were accounted for by the instructorless groups who preferred the classroom experience—the correspondence students showing greater preference for their own group, and the kinescope students tending to show greater preference for the classroom. Subjects' ratings of the course's personal value were highest for the classroom group and lowest for the kinescope group; and they bore a nonsignificant positive relationship to the experimental treatments arranged in the order of most (classroom) to least (correspondence) similarity to customary educational methods.

Four months after the course the correspondence group had retained significantly more than the classroom group in tested achievement, and the classroom group remained significantly higher than the kinescope group in group structure. In addition, the same highly significant positive relationship was found between subjects' preferences among experimental treatments and the treatments to which they had actually been assigned that had appeared throughout the course—the deviates from this trend being distributed between own group and classroom group in the same manner as at the end of the course.

Within the limits imposed upon generalization by the small sample used in this study, these findings are taken to suggest that kinescope or TV techniques are at least as effective as—and independent correspondence study is probably more effective than—conventional class discussion methods for promoting durable factual achievement, alone, in abstract or highly verbal academic subjects. All of the methods, of course, are more effective in this regard than no systematic instruction at all. In addition, the early social mediation of an instructor tends to produce greater familiarity among students in the classroom than in the instructorless kinescope treatment; and students' preferences for their own instructional treatments tend to start low and increase through time under the unfamiliar correspondence method, whereas they tend to decrease with time for the partly familiar kinescope method and remain uniformly high under the customary classroom method.

REFERENCES

1) R. W. Husband, "Television Versus Classroom for Learning General Psychology," *American Psychologist*, 9: 181–183, 1954.

2) R. I. Evans, H. B. Roney, and W. J. McAdams, "An Evaluation of the Effectiveness of Instruction and Audience Reaction to Programming on an Education Television Station," *American Psychologist*, 9: 361–362, 1954. (Abstract).

3) W. C. Olson, *Child Development*. Boston, D. C. Health, 1949.

THE MTAI IN AN AMERICAN MINORITY-GROUP SCHOOL SETTING[1]

I. Differences Between Test Characteristics for Norm and Non-norm Populations

JOSHUA A. FISHMAN

College Entrance Examination Board and The City College, College of the City of New York

Since the publication of the Minnesota Teacher Attitude Inventory (MTAI) (*3*) a number of studies have appeared touching upon the reliability, validity, and factorial composition of the scores obtained upon this instrument (*1, 2, 4, 5, 11, 14*). Future users of this inventory have been counseled to regard it as still in an experimental stage due to the fact that many questions concerning the criteria which the test predicts, the interpretation of its scores, and the stability of its norms remain unanswered. Each large-scale use of this instrument must, therefore, be treated as a research undertaking which attempts to secure partial answers to the above questions in a limited setting. By sharing their insights into the functioning of this instrument, investigators will finally be able to evaluate it properly. The present study is an attempt to investigate certain MTAI test characteristics in a sample drawn from the teacher population of various idealogical-structural types of American Jewish schools in Greater New York City. The score-related characteristics here under study will be (a) type of subject taught, (b) age, (c) sex, (d) amount of post-high school education, and (e) foreign vs. American birth.

(a) In her original instrument Leeds found teachers of "special subjects" (e.g., art, music, etc.) to score lower than the regular classroom teachers (*10, 12*). This finding was also obtained with the normifying population (*3*).

(b) Leeds also found a significant negative relationship between

---

[1] The writer is indebted to Dr. Irving Lorge and Dr. Isidore Chein for their criticisms and suggestions in connection with the original study from which this report is an excerpt, although he alone assumes full responsibility for the data and interpretations here presented.

inventory scores and age (*10*, *12*). However, her collaborators in the normifying study ruled the age factor out statistically by deleting all items which originally showed sharp age differences (*3*).

(c) The MTAI norms for elementary school teachers are based entirely on females, due both to the small percentage of males to be found generally in American public elementary education, as well as to the particular unavailability of males in the grades during wartime. Nevertheless, Leeds mentions "tendencies for male teachers to score lower than females" (*12*).

(d) The MTAI Manual reports that in the normifying population of over seventeen hundred cases the "amount of post-high-school education was significantly and positively related to teacher attitudes in graded elementary and high schools" (*3*).

(e) Finally, in arriving at the norms reported in the MTAI Manual, no foreign-born vs. American-born comparisons were made. Leeds, however, comments that she "also noted a tendency for foreign-born teachers to score somewhat lower than American born" (*12*).

In this study an attempt will be made to trace the relationship between these five test characteristics (as obtained from the normifying population of American midwestern public school teachers) and the MTAI scores of a population largely different from that upon which the norms were constructed. The ideological characteristics of four of the five subgroups into which this population has been divided for the purposes of stratified sampling and data analysis will not be delineated here as this has been done in detail elsewhere (*7*). The fifth subgroup of teachers, namely that labeled "Arts," is composed of teachers employed in all of the four preceding school types. However, rather than being teachers of traditional academic subjects these teachers teach singing, dancing, dramatics, or arts and crafts.

The sample upon which the following data were obtained was randomly selected as part of a larger, unpublished study (*6*) which, in turn, was part of a three-year survey of a score or more aspects of the educational activities of this American minority group.

Due to the non-normality as well as the lack of homoscedasticity of the obtained scores, the data were analyzed by Friedman's two-way nonparametric test of analysis of variance of means by ranks (*9*), rather than by the more searching standard technique.

TABLE I—MEAN MTAI SCORES BY AGE AND SCHOOL TYPE

School Types	Age Groups							
	20–29		30–39		40–49		50–59	
	n	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$
1. OAD	13	60.1	8	48.8	2	6.5	8	20.0*
2. OWD, CWD	18	52.4	6	44.8	17	21.4	5	−10.2
3. ROD	11	62.4†	10	40.8	7	49.8	2	15.0
4. n-rSWD	1	(42.0)	2	31.5	4	19.0	4	9.5*
5. Arts	4	27.8†	5	1.2	1	(30.0)	1	(6.0)
Totals	47	54.6	31	36.6	31	26.8	21	9.2

* Includes the score of one teacher *older* than fifty-nine years.
† Includes the score of one teacher *younger* than twenty years.

1. MTAI Scores by Age Group

The data in Table I unambiguously point toward a highly negative relationship between age and mean MTAI scores. The analysis of variance of means by ranks produces a χ_r^2 value of 10.68 for the data as a whole across all of the school types. As χ_r^2 with 4df equals 9.5 at the 0.05 level, our analysis reveals that the differences between the mean MTAI scores of the various groups are significant at this level.

Cook, Leeds, and Callis, in the MTAI Manual report that in tryout versions of this instrument the correlation between age and score was −0.30. When they divided their sample into two age groups, namely, "below forty" and "forty and above," they found the former to score significantly higher than the latter (at the 0.01 level). These investigators comment that the observed difference is due not only to a constant worsening of teacher attitudes toward children as teacher age increases, but that it is probably also due to the fact that "many older teachers did not have the advantage of the recent emphasis, in teacher training, on educational psychology and on the understanding of child behavior."

As stated before, age was no longer significantly related to scores in the normifying population due to the deletion of items which showed sharp age differences. That age is still significantly related to scores in our *current* sample may be indicative of greater attitudinal differences toward children among the various age groups represented within the population of teachers in schools of this particular American minority group than was the case for the normifying population. The inapplicability of the general norms for teachers of this minority group, at least for more than rough within group comparative purposes, seems clear, at least within the framework of the age factor.

Age itself, however, is usually of sociopsychological interest as an index of types or degrees of exposure. We will see below that age is not as important a variable in itself as it is a concomitant of other dynamic factors imbedded in the matrix of teacher-pupil interaction within the schools of this minority group.

2. MTAI Scores by Sex and American or Foreign Birthplace

The data in Table II point toward a negative relationship between mean MTAI scores and (a) male sex, (b) foreign birthplace. The overall difference between the sexes is somewhat slighter than the differences between the birthplace groups. The analysis of variance of means by ranks again produces a χ_r^2 value of 10.68 which, once more, is significant at the 0.05 level. The differences

TABLE II—MEAN MTAI SCORES BY SEX, BIRTHPLACE AND SCHOOL TYPES

School Types	American-Born				Foreign-Born			
	Male		Female		Male		Female	
	n	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$	n	$\bar{x}$
1	3	16.7	12	69.8	9	17.7	7	42.6
2	21	45.8	11	48.0	9	−2.1	5	10.0
3	13	48.2	14	55.0	2	30.0	1	(17.0)
4	1	(83.0)	1	(42.0)	4	−10.2	5	27.2
5	4	11.0	6	31.8	1	(6.0)	—	—
Totals	42	39.9	44	53.8	25	6.6	18	27.8

Birthplace totals:
American-born, n = 86; $\bar{x}$ = 47.0
Foreign-born, n = 43; x = 15.5
Sex Totals:
Males, n = 67; $\bar{x}$ = 27.5
Females, n = 62; $\bar{x}$ = 46.3

between the mean MTAI scores of the various sex-birthplace groups are thus significant over the data as a whole. We pause, however, to inquire whether the four sex-birthplace combinations, taken *separately*, also differ significantly from *each other*. Utilizing the Link-Wallace method (*13*), described in section 4 below, we find that such *is* the case for the mean score differences between American-born females and foreign-born males or females, as well as between American-born males and foreign-born males. Neither of the within-birthplace sex-differences is significant, however.

In arriving at the norms reported in the MTAI Manual no birthplace-group comparisons were made. It is highly probable, however, that the vast majority of American midwestern public school teachers upon whom this instrument was standardized were American-born.[2]

The significant differences between the birthplace-groups can probably be ascribed to differences in philosophy of education encountered in teacher training, as well as to differences in the psychological atmosphere of the family, school, and total adult-child interaction during the teacher's own childhood and youth. In addition, the birthplace category interacts appreciably with the age variable, as the foreign-born teachers are, generally, older than the American-born teachers. The average age of the former is forty-five years, whereas that of the latter is thirty-two years.

The sex difference in mean MTAI score is an interesting although not entirely unexpected finding. Many boards of education throughout the country do not engage males to teach in the lower elementary grades,[3] notwithstanding the shortage of teachers. This policy is due to a conviction that females are better suited to the teaching of young children. In addition to some possible difference in the degree to which males and females may understandingly empathize with young children, there are probably also some psychological needs of young children which females can better meet than males. That any sex differences which may exist in teacher attitudes toward young children cannot be very great among American public school teachers in general is attested to by the fact that Cook, Leeds and Callis report that "men and women graduate students (majoring in education, with an average of ten years' teaching experience) . . . have mean MTAI scores which are not significantly different" (*3*).

---

[2] The preponderantly middle-western nature of the normifying population also allows us to deduce the modal nature of other sociopsychological characteristics which differentiate between it and the current sample.

[3] In New York City, e.g., males are not engaged to teach below the fifth grade.

The significant difference between mean MTAI scores of males and females in our minority group sample may be explained by factors additional to those mentioned above. The males in our sample are both older (males: average age = forty; females: average age = thirty-three) and to a larger extent foreign-born (males: thirty-seven per cent foreign-born; females: twenty-nine per cent foreign-born). In addition, traditional Jewish education for males has been more demanding and more subject-centered than that for females. To some extent current teachers of this minority group may be reacting to their pupils and to their rôles as teachers in the light of their own childhood and adolescent experiences with Jewish education and with the teacher-pupil rôles considered appropriate in those times.

These sex and birthplace differences in mean score are not provided for in the general norms. Some such provision is obviously needed for interpreting scores of samples drawn from the universe of this minority group's teachers.

3. MTAI Scores by Extent of Secular Education

The data in Table III clearly indicate a positive relationship between extent of general secular (as opposed to minority-group sponsored) education and mean MTAI scores. The analysis of variance of means by ranks produces a χ_r^2 value of 9.96. Since χ_r^2 with 4df is 9.5 at the 0.05 level, the over-all differences between the mean MTAI scores of the various secular educational levels are significant at this level.

As stated before, amount of higher education was found to be positively and significantly related to MTAI scores in the normifying population of seventeen hundred cases.

In our current sample the effect of increased secular education, as might be expected, interacts, to some extent, with the age and birthplace findings discussed previously. The older, European-born and -trained teachers were also found to have received less formal education of a general secular variety than have the younger, American-born teachers (*6*). However, there is also much to be

TABLE III—MEAN MTAI SCORES BY EXTENT OF GENERAL SECULAR EDUCATION

School Types	Extent of Secular Education							
	NC*		C		B		M	
	n	x̄	n	x̄	n	x̄	n	x̄
1	3	12.3	9	30.7	14	43.1	5	73.0
2	2	−10.0	16	5.8	20	49.0	8	58.4†
3	—	—	5	42.8	13	38.1	12	61.3†
4	3	−17.5	6	45.7	—	—	2	−10.0
5	2	10.0	4	15.0	1	(−28.0)	4	35.7
Totals	10	2.8	40	22.9	48	42.7	31	53.4

* NC = no college; C = college attended, no degree; B = bachelor's degree; M = master's degree.

† Includes the score of one Ph.D.

said in favor of an appreciable effect for a straight "secular education" factor. Increased secular education most frequently signifies increased familiarity with concepts of modern education, educational psychology, child development, mental hygiene, etc. On the whole, an increased awareness and concern for the emotional rights and needs of children may be expected to parallel increased exposure to modern views in these subject areas.

In this respect, then, the MTAI is functioning for this non-norm group in much the same manner as it was intended to function for samples much more similar to the normifying population.

Increased secular education may be related to mean MTAI scores in yet another way, namely, via increased test sophistication. As a fakable instrument (*14*) the MTAI would be most vulnerable to faking by those most experienced with modern objective tests, i.e., by those groups with the greatest amount of general secular education. To the extent that this relationship between faking of test response and higher educational attainment is present in our data (as well as in the norm data), to that (unknown) extent are our findings suspect. All that can be said in the current connection is that respondent-anonymity was permitted to encourage complete frankness in the filling out of this questionnaire.

TABLE IV—MEAN MTAI SCORES OF TWO SUCCESSIVE SAMPLES OF TEACHERS IN FIVE TYPES OF SCHOOLS

School Types	First Sample			Second Sample			Total		
	n	x̄	S.D.	n	x̄	S.D.	n	x̄	S.D.
1	26	43.9	45.0	5	45.8	34.5	31	44.2	43.4
2	41	33.6	47.8	5	28.4	53.5	46	33.0	48.4
3	5	47.6	40.2	5	56.6	25.7	30	49.1	63.5
4	2	64.0	22.0	9	10.1	41.8	11	19.9	44.2
5	8	11.0	50.6	1	21.7	20.0	11	13.9	44.7
Totals	102	40.2	44.2	27	33.1	40.3	129	36.1	45.5

4. MTAI Scores and Subject Matter ("Academic" vs. "Arts") Taught

From all of the foregoing tables, and especially from Table IV which follows, it is obvious that the teachers in group 5, i.e., the teachers of art, music, dancing, dramatics, or arts and crafts, obtained a lower mean MTAI score than did any other group of teachers. This finding is in agreement with the findings reported by Leeds (*10*, *12*) and with the information given in the Manual (*3*). At this point, however, we pause to carry the analysis one point further and ask whether the mean score of the "arts" teachers is significantly different from those of the means obtained by the other teachers.

A somewhat roundabout attack at this problem is necessary, due to the regrettably unsatisfactory nature of the reply which analysis of variance designs give to questions of the type: "Is column (or row) X significantly different from column (or row) Y or Z?" (*13*).

We will, therefore, first enquire whether the five groups of teachers differ significantly over the data as a whole. If the answer to this question is in the affirmative, we will then enquire as to which particular groups differ in total mean scores from the others. In this roundabout manner we can determine from which, if any, of the groups of teachers of academic subjects the arts teachers might differ significantly.

To answer these questions we will utilize the data of Table IV. These data were originally collected in order to determine whether additional sampling was necessary before embarking on the terminal analyses of the category comparisons reported in sections 1 to 3 above.

The data in Table IV indicate a highly consistent rank order across the two samples for the mean MTAI scores of the five groups of teachers. The analysis of variance of means by ranks produces a χ_r^2 value of 32.8. Since χ_r^2 with 4df is 13.3 at the .01 level, the over-all differences between the mean MTAI scores of the various types of teachers may be considered highly significant. This being the case, we may now proceed to enquire whether the mean score of the teachers in group 5 departs significantly from those of the other four groups.

Mosteller and Bush (*13*, pp. 305–307) present a nonparametric method, first proposed by R. F. Link and D. L. Wallace, for tackling such problems in two-way analyses of variance with one observation in each cell. This method employs the range as a measure of variation rather than sums of squares. It has here been employed after converting means to ranks. The "allowance for column totals" obtained when employing this method turns out to be so large, that we are justified in "recognizing" a difference between the total mean score of arts teachers and the total mean of only one other teaching group, namely group 3. The obtained size of the "allowance" is large, due to the small number of rows and columns employed in our analysis. Thus we are forced to conclude that the mean score of the "arts" and other "special" teachers, although low indeed, is not significantly different than the mean scores of most of the various types of teachers of academic subjects.[4] In this respect, therefore, we may conclude that although the MTAI functions on this non-norm group similarly to its functioning on the norm population, the reliability of this phenomenon is not one which we may accept with complete confidence.

SUMMARY AND CONCLUSION

When the MTAI was administered to teachers from a non-norm universe, a number of the characteristics which this instrument reveals with American midwestern public school teachers no longer obtained, while other instrument characteristics continued to obtain.

---

[4] Only nonsignificant differences resulted from applying to our data the well-known Tukey (*15*) test for significance of gap after performing a common analysis of variance with one observation in each cell.

Those characteristics of the MTAI which obtain for the norm and norm-similar populations, but which *did not obtain* for this sample from a universe of teachers in schools of an American minority group were: (a) the constancy of MTAI scores over increased age.

Those characteristics of the MTAI which obtain for the norm and norm-similar populations, and which *did also obtain* for this sample from a universe of teachers in schools of an American minority group were: (a) the lower scores for teachers of "special subjects"; (b) the higher scores for teachers with greater amounts of post-high school education.

In addition, it was discovered that mean MTAI scores varied significantly with two other independent variables not employed in the MTAI normifying studies: (a) sex, males scoring significantly lower than females, (b) birthplace, foreign-born teachers scoring significantly lower than American-born teachers.

Although the teachers of all school types obtained relatively low MTAI scores when compared with the norm of American public school teachers, the highest over-all mean scores were obtained for school types in which young, American-born, females with advance college degrees predominate on the teaching staffs (type 1 and, particularly, type 3). Conversely, those school types in which most teachers are old, foreign-born, males with little formal general (secular) education obtained the lowest over-all mean scores.[5]

Although departing somewhat from its functional and structural characteristics as a measure of teacher attitudes toward children in the norm-population, the MTAI scores in this non-norm population still varied meaningfully with independent variables affecting teacher attitudes and classroom performance.

A final question which remains as we bring these comments to a close is: does the MTAI continue to measure degree of teacher orientation toward and acceptance of the child's emotional and developmental needs when it is administered in such a non-norm universe? A partial answer to this question will be found in a subsequent publication (*8*).

---

[5] In a further study it was demonstrated that of these four variables, extent of secular education was found to be the most significantly related to MTAI scores (the three other variables being held constant), followed by sex, as the second most significant variable. The contributions of age and, particularly of nativity, were negligible (*6*).

REFERENCES

1) Robert Callis, "The Efficiency of the Minnesota Teacher Attitude Inventory for Predicting Interpersonal Relations in the Classroom," *Journal of Applied Psychology*, 37: 82–85, 1953.

2) R. Callis and J. L. Ferguson, "Cluster Analysis of the Minnesota Teacher Attitude Inventory," Technical Report No. 2 for Contract NONR 649 (00), University of Missouri, 1953.

3) W. W. Cook, R. Callis, and C. H. Leeds, *Minnesota Teacher Attitude Inventory*. New York, Psychological Corporation, 1951.

4) N. M. Downie and C. R. Bell, "The Minnesota Teacher Attitude Inventory as an Aid in the Selection of Teachers," *Journal of Educational Research*, 46: 699–704, 1953.

5) J. L. Ferguson, "A Factorial Study of the Minnesota Teacher Attitude Inventory," *Dissertation Abstracts*, 13: 1087, 1953.

6) J. A. Fishman, "A Study of the Attitudes Held by Teachers in Jewish Schools toward Children and toward Various Aspects of Their Own Work." New York, Jewish Education Committee of New York, 1953.

7) J. A. Fishman, "Negative Stereotypes Concerning Americans Among American-born Children Receiving Various Types of Minority-group Education," *Genetic Psychology Monographs*, 51: 107–182, 1955.

8) J. A. Fishman, "The MTAI in an American Minority-group School Setting: II. Validation as a Test of Pupil-directedness" (in preparation).

9) M. Friedman, "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance," *Journal of the American Statistical Association*, 32: 675–701, 1937.

10) C. H. Leeds, "A Scale for Measuring Teacher-Pupil Attitudes and Teacher-Pupil Rapport," *Psychological Monographs*, 64: 1950.

11) C. H. Leeds, "A Second Validity Study of the Minnesota Teacher Attitude Inventory," *Elementary School Journal*, 52: 398–405, 1952.

12) C. H. Leeds, Personal communication, dated June 29, 1953.

13) F. Mosteller and R. B. Bush, "Selected Quantitative Techniques," in *Handbook of Social Psychology*, edited by G. Lindzey. Addison-Wesley, Cambridge, 1954.

14) W. Rabinowitz, "The Influence of Instructional Sets on Minnesota Teacher Attitude Inventory scores," New York, Office of Research and Evaluation Division of Teacher Education, City College of New York, 1953.

15) J. W. Tukey, "Comparing Individual Means in the Analysis of Variance. "*Biometrics*. 5: 99–114, 1949.

KINDERGARTEN TRAINING AND GRADE I READING[1]

IRENE FAST[2]

University of Michigan

The influence of kindergarten training on the later development of the child has been a research problem since the beginning of the century when kindergartens were becoming increasingly widespread on this continent. The results of investigations have suggested that children with kindergarten training advance more rapidly in reading (*4*, *5*, *7*) and arithmetic (*1*), and are generally more advanced throughout the elementary school (*2*, *6*). Despite the consistency of these findings, however, the question remains open because of the difficulty of conducting systematically controlled research in this area.

Specifically, it has been impossible to obtain comparable groups of children with and without kindergarten training. Where attendance at kindergarten is voluntary, as it apparently was in all studies, uncontrolled differences in home background may exist between children who attend kindergarten and those whose parents neglect, refuse, or are for some reason unable to send them. Differences in age of entrance into grade I exist in at least one case (*2*), and may be present in others where no information is given. Finally, where children with kindergarten training and those without have not attended the same schools, as seems to be the case in the MacLatchy (*5*) and Brueckner (*2*) studies, differences in educational policy and administration may have influenced the results to an unknown extent.

It seems clear that until investigations controlling these factors can be made no final conclusions can be drawn. Fortunately the present writer was able to make use of an existing situation in which age, home background, and school environment of the subjects could be controlled. In the three urban schools in which the study was conducted children whose fifth birthdays fell in

---

[1] This investigation was supported by Canadian Federal Health Project 605-5-147.

[2] The writer gratefully acknowledges the assistance of Dr. R. L. Cutler in the preparation of this manuscript for publication.

November and December of the current school year (1953) were excluded from kindergarten training because accommodation was limited. The following year they were admitted to grade I and placed in the same classrooms as first grade children who had received kindergarten training. Thus it was possible to examine differences between one hundred and thirty-four children with kindergarten training and forty-six children who had no opportunity for such training. No difference between the two groups other than age was introduced by the criterion of selection for kindergarten training.

It was decided to focus the study on reading progress in the first grade, both because reading is generally considered to be of primary importance in grade I, and because standardized tests of reading are readily available. It was predicted that initial reading scores of children with kindergarten training would be higher than scores of children without such training, and that this advantage would be maintained throughout the school year.

METHOD

Subjects. All grade I children enrolled in the three urban schools were given a battery of tests. The test results of one hundred and thirty-four children with kindergarten training and forty-six children without such training were utilized in the present study. The test results of children who were repeating grade I and of those absent during one or more of the testing sessions were not used.

Materials and Procedure. Following is a list of tests included in the battery administered to the above-described population:

(1) The Dominion Group Test of Reading Readiness, Short Form, Omnibus Type B was administered to all grade I children in the three schools four weeks after the beginning of the fall term. It was given at this time in order that the children would have had some opportunity to become accustomed to the classroom situation without having had a great deal of formal reading instruction.

(2) The Survey Test in Silent Reading (First Experimental Edition, now being standardized) was administered at the beginning of February to discover whether differences between the kindergarteners and nonkindergarteners were still evident. The

two parts, Word Recognition and Paragraph Reading, were treated separately in the comparison of groups.

(3) The Achievement Tests in Silent Reading; Primary, Grade I, type III; Paragraph Reading Test; and The Group Test of Learning Capacity, Primary, were administered in May.

RESULTS

The initial requirement in presenting the results of this study is to demonstrate that no differences in age or home background influenced the observed differences in reading scores.

Seven of the one hundred and thirty-four kindergarteners had birthdates in November and December and had received kindergarten training in other schools. There seemed to be no *a priori* reason for excluding these from the sample. Twenty-one of the forty-six nonkindergarteners' birthdates were between January and October. These children had not been excluded from kindergarten but had failed to attend, probably because their parents had been unable to arrange for transportation and they lived some distance from school (maximum five miles). To investigate the possibility of differences in reading ability or I.Q. between these children and the nonkindergarteners whose birthdates fell in November and December t tests were made. No significant difference in either reading or I.Q. was found. Therefore the forty-six nonkindergarteners are treated as one group throughout the study.

No difference between kindergarteners and nonkindergarteners in home environment was introduced by the criterion of selection for kindergarten training. To explore the possibility that differences in economic status existed between the groups, the occupations of the fathers of the subjects, where they were available, were plotted on a six-point scale published in the 1947 census data (*3*). A chi-square test showed no significant difference in economic status between fathers of kindergarteners and nonkindergarteners.

Since only children who were five years of age before November 1, 1953, were admitted to kindergarten, and most nonkindergarteners reached age five in November or December, 1953, the kindergarteners were slightly older than the nonkindergarteners. Scores of kindergarteners and nonkindergarteners on all tests were plotted against age. Because no relation between scores and age within the one year period which included all dates of birth

TABLE I—THE MEANS ON FOUR TESTS AND SIGNIFICANCE OF DIFFERENCES BETWEEN KINDERGARTENERS AND NONKINDERGARTENERS EQUATED ON I.Q. AND M.A.

Month of Administration	Test	Mean Scores		Subjects Equated on	
		K	Non-K	I.Q.	M.A.
October	Reading readiness	8.67	4.03	0.0025	0.0025
February	Word recognition	13.71	11.83	0.028	0.013
February	Paragraph reading	5.45	3.96	0.018	0.005
May	Paragraph reading	13.30	9.54	0.0015	0.035

was found, it was concluded that age difference within this period did not significantly influence reading scores.

Differences in I.Q. (scores on the Dominion Group Test of Learning Capacity, Primary) were controlled by matching kindergarteners and nonkindergarteners. The means of reading scores in each interval of five I.Q. points were computed for each group on each reading test. The mean scores of kindergarteners and nonkindergarteners were compared using a one-tailed t test for matched individuals. The significance of differences is shown in Table I.

To control both age and I.Q., the groups were then equated as described above, using five-month mental age intervals. As is shown in Table I the groups differed significantly on each test. The similarity between results when children were equated on I.Q. and M.A. scores corroborates the finding that age differences within and between groups did not significantly affect the reading scores.

DISCUSSION

In general the results of previous studies (*4*, *5*, *7*) have been confirmed: kindergarteners achieve significantly higher scores in reading tests than nonkindergarteners. At least three further considerations determine the value of the results.

The first concern is the generality of the results beyond the schools in which the study was conducted. The three schools were within the public education system. Because a chief area of variability within authorized programs concerns formal reading in kindergarten it should be mentioned that, while explicit reading readiness programs were followed, reading was not formally taught in the kindergartens of any of the three schools. The use of widely distributed, standardized reading series in the first grade suggests that reading experiences were broadly similar to those found throughout this continent. The quality of instruction was not unduly influenced by any single teacher because the children were distributed among three teachers in kindergarten and among nine in the first grade. These factors suggest that similar results would be obtained if the study were repeated elsewhere.

The other two questions concern the extent to which similar differences may be found in other aspects of development and the length of time in the academic progress of the children that differences in achievement persist. At present these questions cannot be answered decisively. If the kindergarten reading readiness program is most closely related to reading progress in grade I we would not expect to find differences between kindergarteners and nonkindergarteners in other aspects of development. However, the major part of the kindergarten program is devoted to more general development in social skills, the ability to work independently and in groups, and the acquisition of verbal and manual skills. If these are the more important factors differences between kindergarteners and nonkindergarteners in subjects other than reading might be found in grade II or later.

It is hoped that these questions may be clarified by an extension of the study during the current school year.

SUMMARY

Reading tests were administered at the beginning, middle, and end of the grade I year to one hundred and thirty-four children who had spent a year in kindergarten and forty-six otherwise comparable children who had not. Group tests of mental capacity were administered at the end of grade I.

Children with kindergarten training were found to achieve significantly higher scores on all reading tests than children without such training. The possibility of clarifying the degree to which such differences existed in other academic areas and persisted in later years was explored.

REFERENCES

1) Y. Bergamini and W. Swanson, "Does Kindergarten Make a Difference?" *School Executive*, 74: 54–55, Dec., 1954.

2) L. J. Brueckner and others, *The Changing Elementary School, The regents' inquiry*, 80–81. New York, Inor Publishing Co., 1939.

3) *Canadian Census Occupational Data for Counsellors*. Ottawa, 1947.

4) W. L. Gard, "The Influence of Kindergarten on Achievement in Reading," *Educational Research Bulletin*, 3: 135–138, 1924.

5) J. MacLatchy, "The Influence of Kindergarten Attendance on Progress in Grade I," *Educational Research Bulletin*, 3: 361, 1924.

6) W. J. Peters, "The Progress of Kindergarten Pupils in Elementary Grades," *Journal of Educational Research*, 7: 117–126, 1923.

7) W. E. Pratt, ' A Study of the Differences in the Prediction of Reading Success of Kindergarten Children and Non-kindergarten Children, *Journal of Educational Research*, 42: 525–533, 1949.

BOOK REVIEWS

ARTHUR E. FINK, EVERETT E. WILSON AND MERRILL B. CONOVER. *The Field of Social Work*, third edition. New York: Henry Holt and Company, pp. 630, 1955.

An already good text has been greatly improved in the third edition. Concerning the addition of two authors, the Preface says: "One evidence of the acceleration as well as of the range within the social work field is reflected in the triple authorship of this revision . . . a volume that undertakes to deal with as extensive an area as social work now demands the thinking and the experience of other persons."

The work is organized about sixteen main heads: the problems that people present; the development of social services—the European background; social services in America—from the almshouse to social security; social services in America—from the church to the charity organization society movement; social casework; social services in a family-focused agency; services in a local welfare department; welfare services for children; psychiatric social work; medical social work; the correctional services; school social work; social services for the aged; social group work; community organization for social welfare; and the profession of social work. The chapters on the problems that people bring to social agencies, on social casework, and on social services to the aged are new. New illustrative material has been added and the bibliographies have been brought up to date. Sources for films are indicated.

The historical chapters give a valuable background for the understanding of present-day social work. The complexity of modern social life is reflected in the greater complexity and difficulties of social work. The authors bring out the development or evolution of social services, relating the various steps to the fundamental sciences upon which they are dependent,—medical, psychiatric, psychological, psychoanalytic, and the relations of education, law and correctional procedures, etc., to the work as a whole.

The development of social work as a profession is discussed, including a platform and an association, organizations, training, research, jobs, salaries, publications, etc. The new profession is not something which is to be taken up casually by those who have failed in some other work. It is something which now stands on its own feet and is to be respected by those who understand it. Its interrelations with many organizations and funds are clearly indicated in this volume.

For this tightly packed and unpadded book the present reviewer has little of adverse criticism. But although prevention is mentioned and in several places implied, it seems that more on this most important subject might have found place; it does not appear in the Index. Mental Hygiene finds some place but not the more modern development under the now more common head of Mental Health. This most important problem of civilization might have had a little more attention. There is, however, considerable under Mental Illness.

Case studies at the ends of chapters make a valuable contribution to the understanding of the social work problems. There are some footnotes, and especially excellent bibliographies at the end of chapters.

The book will be useful not only as a text but also as a reference book for many details in the development of social work.

A. S. Edwards

The University of Georgia

PUBLICATIONS RECEIVED

FRANCES LAUGHLIN. *The Peer Status of Sixth and Seventh Grade Children.* New York: Bureau of Publications, Teachers College, Columbia University, 1954, pp. 85. $2.75.

PAUL F. LAZARSFELD AND MORRIS ROSENBERG, Editors. *The Language of Social Research.* Glencoe, Ill.: The Free Press, 1955, pp. 590. $6.75.

GEORGE F. J. LEHNER AND ELLA KUBE. *The Dynamics of Personal Adjustment.* New York: Prentice-Hall Inc., 1955, pp. 498. $5.25.

PAUL V. LEMKAU. *Mental Hygiene in Public Health.* New York: McGraw-Hill Book Company, Inc., 1955, pp. 486. $8.00.

CLARENCE LEUBA. *The Sexual Nature of Man.* Garden City, New York: Doubleday and Company, Inc., 1954, pp. 40 (paper). $0.85.

HILDA LEWIS. *Deprived Children.* New York: Oxford University Press, 1954, pp. 163. $1.55.

FRANK A. LOGAN, DAVID L. OLMSTED, BURTON S. ROSNER, RICHARD D. SCHWARTZ AND CARL M. STEVENS. Behavior Theory and Social Science. New Haven: Yale University Press, 1955, pp. 188. $3.00.

ROBERT E. MAINER. *Attitude Change in Intergroup Education Programs.* Lafayette, Indiana: Purdue University, Division of Educational Reference, 1954, pp. 56 (paper).

DONALD R. MATTHEWS. *The Social Background of Political Decision-Makers.* Garden City, New York: Doubleday & Company, Inc., 1954, pp. 71 (paper). $0.95.

W. F. MCAULEY. *The Concept of Schizophrenia.* New York: Philosophical Library, 1954, pp. 145.

MARGARET G. MCKIM. *Guiding Growth in Reading in the Modern Elementary School.* New York: The Macmillan Co., 1955, pp. 528. $5.25.

MARGERY D. MCMULLIN. *How to Help the Shut-in Child.* New York: E. P. Dutton & Company, Inc., 1954, pp. 192. $2.75.

QUINN MCNAMAR. *Psychological Statistics.* Second Edition. New York: John Wiley & Sons, Inc., 1955, pp. 408. $6.00.

MARGARET MEAD AND MARTHA WOLFENSTEIN, Editors. *Childhood in Contemporary Cultures.* Chicago: The University of Chicago Press, 1955, pp. 473. $7.50.

Angela Medici. *L'Ecole et L'Enfant.* Paris: Presses Universitaires De France, 1955, pp. 148.

Paul E. Meehl. *Clinical vs. Statistical Prediction: A Theoretical Analysis and A Review of the Evidence.* Minneapolis: University of Minnesota Press, 1954, pp. 149. $3.00.

Max Meenes. *Studying and Learning.* Garden City, New York: Doubleday & Company, Inc., 1954, pp. 68 (paper). $0.95.

John U. Michaelis. *The Prediction of Success in Student-teaching from Personality and Attitude Inventories.* Berkeley: University of California Press, 1954, pp. 64 (paper). $0.85.

Cecil V. Millard. *School and Child: A Case History.* East Lansing: The Michigan State College Press, 1954, pp. 221. $3.75.

Harold E. Moore and Newell B. Walters. *Personnel Administration in Education.* New York: Harper & Brothers, 1955, pp. 476. $5.00.

Douglas M. More. *Developmental Concordance and Discordance During Puberty and Early Adolescence.* Champaign, Ill.: Child Development Publications, 1955, pp. 128.

Kate Hevner Mueller. *Educating Women for a Changing World.* Minneapolis: University of Minnesota Press, 1954, pp. 302. $4.75.

1954 Invitational Conference on Testing Problems. Princeton, N. J.: Educational Testing Service, 1954, pp. 134 (paper).

Elizabeth K. Nottingham. *Religion and Society.* Garden City, New York: Doubleday & Company, Inc., 1954, pp. 84 (paper). $0.95.

Yoshikazu Ohwaki. *Tohoku Psychologica Folia.* Sendai, Japan: Tohoku University, 1955, pp. 158 (paper).

Eamon N. M. O'Sullivan. *Textbook of Occupational Therapy.* New York: Philosophical Library, Inc., 1955, pp. 319. $10.00.

A. Harry Passow, Miriam Goldberg, Abraham J. Tannenbaum and Will French. *Planning for Talented Youth.* New York: Bureau of Publications, Teachers College, Columbia University, 1955, pp. 84 (paper). $1.00.

A. Harry Passow, Matthew B. Miles, Stephen M. Corey and Dale C. Draper. *Training Curriculum Leaders for Coöperative Research.* New York: Bureau of Publications, Teachers College, Columbia University, 1955, pp. 158 (paper). $1.35.

Catharine Patrick. *What is Creative Thinking?* New York: Philosophical Library, 1955, pp. 210. $3.00.

JON S. PETERS, Editor. *Trends in School Planning*. Stanford, California: Stanford University Press, 1955, pp. 119 (paper). $4.00.

ELEANOR M. PETERSON. *Aspects of Readability in the Social Studies*. New York: Bureau of Publications, Teachers College, Columbia University, 1954, pp. 118. $3.50.

TRUMAN M. PIERCE, JAMES B. KINCHELOE, R. EDGAR MOORE, GALEN N. DREWRY AND BENNIE E. CARMICHAEL. *White and Negro Schools in the South*. New York: Prentice-Hall Inc., 1955, pp. 338. $4.95.

HENRI PIERON, MAURICE REUCHLIN, R. BIZE, COLETTE BENASSY-CHAUFFARD, S. PACAUD AND P. RENNES. *L'Utilisation des Aptitudes, Livre III*. Paris: Presses Universitaires de France, 1954, pp. 759 (paper).

KENNETH L. PIKE. *Language in Relation to a Unified Theory of the Structure of Human Behavior*. Glendale, California: Summer Institute of Linguistics, 1954, pp. 170 (paper).

L. A. COSTA PINTO. *O. Balconista, Analise Das Profissoes Comerciais, Monografia N. 2*. Departamento Nacional Do Senac: Divisao De Orgaos Regionais, 1954, pp. 62 (paper).

EDWARD PODOLSKY, Editor. *Management of Addictions*. New York: Philosophical Library, 1955, pp. 413. $7.50.

RALPH C. PRESTON, Editor. *Teaching World Understanding*. New York: Prentice-Hall Inc., 1955, pp. 207. $3.95.

Promising Curriculum Practices in Secondary Schools in Indiana, Bulletin of the School of Education, Indiana University. Indiana University: Division of Research and Field Services, 1954, pp. 98 (paper).

ROBERT REDFIELD. *The Educational Experience*. Pasadena, California: The Fund for Adult Education, 1955, pp. 61 (paper).

FRANCIS REITMAN. *Insanity, Art, and Culture*. New York: Philosophical Lirary, 1954, pp. 111. $3.75.

MARGARET A. RIBBLE. *The Personality of the Young Child*. New York: Columbia University Press, 1955, pp. 126. $2.75.

A. A. ROBACK, Editor. *Present-Day Psychology*. New York: Philosophical Library, 1955, pp. 995. $12.00.

RAGNAR ROMMETVEIT. *Social Norms and Rôles. Explorations in the Psychology of Enduring Social Pressure*. Minneapolis: University of Minnesota Press, 1955, pp. 167. $2.50.

Leon Roth. *Jewish Thought as a Factor in Civilization.* Paris: UNESCO, 1954 (paper).

Harold Rugg and William Withers. *Social Foundations of Education.* New York: Prentice-Hall, Inc., 1955, pp. 771. $5.40.

Harry Ruja. *Psychology for Life.* New York: McGraw-Hill Book Company, Inc., 1955, pp. 427. $4.75.

Dagobert D. Runes. *Treasury of Philosophy.* New York: Philosophical Library, 1955, pp. 1280. $15.00.

Robert R. Rush and Frederick Mosteller. *Stochastic Models for Learning.* New York: John Wiley & Sons, Inc., 1955, pp. 365. $9.00.

David H. Russell. *The Dimensions of Children's Meaning Vocabularies in Grades Four Through Twelve.* Berkeley and Los Angeles: University of California Press, 1954, pp. 414 (paper). $1.25.

Roy Schafer. *Psychoanalytic Intrpretation in Rorschach Testing: Theory and Application.* New York: Grune & Stratton, 1954, pp. 446. $8.75.

C. Winfield Scott and Clyde M. Hill, Editors. *Public Education Under Criticism.* New York: Prentice-Hall, Inc., 1954, pp. 414. $4.75.

Jerome M. Seidman. *Readings in Educational Psychology.* Boston: Houghton Mifflin Company, 1955, pp. 402 (paper). $3.25.

Services for the Physically Handicapped. United Nations, Department of Social Affairs, 1954, pp. 31 (paper).

Hirsch Lazaar Silverman. *Education Through Psychology. A Subjectivist-Idealist Format for Education.* New York: Exposition Press, 1954, pp. 58. $3.00.

Peter Slade. *Child Drama.* New York: Philosophical Library, 1955, pp. 379. $10.00.

William Sloane. *Children's Books in England and America in the Seventeenth Century.* New York: King's Crown Press, Columbia University, 1955, pp. 251.

Speak Truth to Power, A Quaker Search for an Alternative to Violence. A Study of International Conflict. Prepared for the American Friends Service Committee. pp. 71 (paper).

George D. Spindler, Editor. *Education and Anthropology.* Stanford, California: Stanford University Press, 1955, pp. 302 (paper). $5.50.

Herbert F. Spitzer. *Spitzer Study Skills Test.* Yonkers-on-Hudson, N. Y.: World Book Company (paper).

Jean Stoetzel. *Without the Chrysanthemum and the Sword, A Study of the Attitudes of Youth in Post-War Japan.* New York: Columbia University Press, 1955, pp. 334. $4.00.

Leland H. Stott. *The Longitudinal Study of Individual Development.* Detroit: The Merrill-Palmer School, 1955, pp. 115 (paper).

Ruth Strang, Constance M. McCullough and Arthur E. Traxler. *Problems in the Improvement of Reading.* New York: McGraw-Hill Book Company, Inc., 1955, pp. 426. $5.00.

Strengthening Family Life Education in our Schools: Report of the Midwest Project on In-service Education of Teachers. A Project of The American Social Hygiene Association, pp. 185 (paper). $1.00.

Edward K. Strong, Jr. *Vocational Interests 18 Years After College.* Minneapolis: University of Minnesota Press, 1955, pp. 207. $3.75.

Studies in Education. 7. The Bearings of Recent Advances in Psychology on Educational Problems. University of London Institute of Education. London: Evans Brothers, Ltd., 1955, pp. 215 (paper).

August Pi Suner. *Classics of Biology.* New York: Philosophical Library, 1955, pp. 337. $7.50.

Donald E. Super. *Opportunities in Psychology.* New York: Vocational Guidance Manuals, 1955, pp. 96 (paper).

Ernst W. Swanson and John A. Griffin, Editors. *Public Education in the South Today and Tomorrow.* Chapel Hill: The University of North Carolina Press, 1955, pp. 137. $5.00.

Merle W. Tate. *Statistics in Education.* New York: The Macmillan Company, 1955, pp. 597. $6.00.

Teaching Spelling, Course of Study and Manual. Curriculum Bulletin, 1953–1954 Series, Number 6. Board of Education of the City of New York, 1954, pp. 135 (paper).

Robert L. Thorndike and Elizabeth Hagen. *Measurement and Evaluation in Psychology and Education.* New York: John Wiley & Sons, Inc., 1955, pp. 575. $5.50.

R. M. Thrall, C. H. Coombs, and R. L. Davis, Editors. *Decision Processes.* New York: John Wiley & Sons, Inc., 1954, pp. 332. $5.00.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 48 February, 1957 Number 2

A SECOND QUARTER CENTURY OF DELAYED RECALL OR RELEARNING AT 80

L. B. HILL

Fifty years ago, A. E. Rejall and L. B. Hill (the present writer, referred to later in this study as the experimenter), then students at Teachers' College, Columbia University, learned to typewrite with a certain degree of proficiency, under experimental conditions. The experiment consisted of learning to typewrite by writing a one hundred-word paragraph, and by copying, at the same sitting, a page of new material containing three hundred words. The original data were analyzed by the late Dr. Edward L. Thorndike and published in the *Pedagogical Seminary* of December, 1913, under the title "Practice in the Case of Typewriting."[1]

Twenty-five years later, the writer, who had had very little significant practice in using the typewriter in the meantime, recalled the learning formerly acquired. The method of recall was the same as that used in the earlier experiment. The data thus secured were published in the *Pedagogical Seminary and Journal of Genetic Psychology* under the title "A Quarter Century of Delayed Recall."[2]

Now, twenty-five years after the first recall, and fifty years after the first learning, the writer presents herewith certain data from a second quarter century of delayed recall, together with his own interpretation and comments upon the findings.

Certain facts, more or less significant to a second recall, should be kept in mind, such as the age of the learner and the fact that he retired from teaching ten years ago. Since retirement he has done no significant work of a professional nature. He has changed en-

---

[1] Hill, L. B., A. E. Rejall, and E. L. Thorndike. Practice in the Case of Typewriting. *Ped. Sem.*, 1913, 20, 315–529.

Hill, L. B. A Quarter Century of Delayed Recall. *Ped. Sem. and Jo. Gen. Psy*. 1934, 44, 231–238.

Bureau Ednl. Research
COLLEGE
Dated
Acs No

vironment and has been identified with some civic organizations in the community in which he has lived. He has engaged in some pleasurable activities such as taking care of a new home, planting flowers and shrubbery, and learning to play shuffleboard, to dance, and to play a better game of bridge. His physical condition may be described as good. The only real physical impairment which he recognizes is in eyesight which makes his reading a bit slow and difficult.

The second recall of learning to typewrite was undertaken (1) because of a persistent interest in the learning process—an interest which lends pleasure in learning anything new, even at the expense of considerable effort. (2) Since writing with a pen has become a little difficult, it will be worthwhile and pleasurable to be able to use the typewriter. (3) It will also be interesting to find out whether the experimenter, at his present age, can still learn, and to see what factors enter into the process. Other factors favorable to undertaking the project may be mentioned as a feeling of well-being, living in an environment where a great many old people learn new hobbies, play a great deal, and enjoy life in retirement.

Reference to Figure 1 shows that in the first experiment the learner wrote the one hundred-word paragraph at the rate of 7.4 words per minute at the first sitting. (In all three experiments two

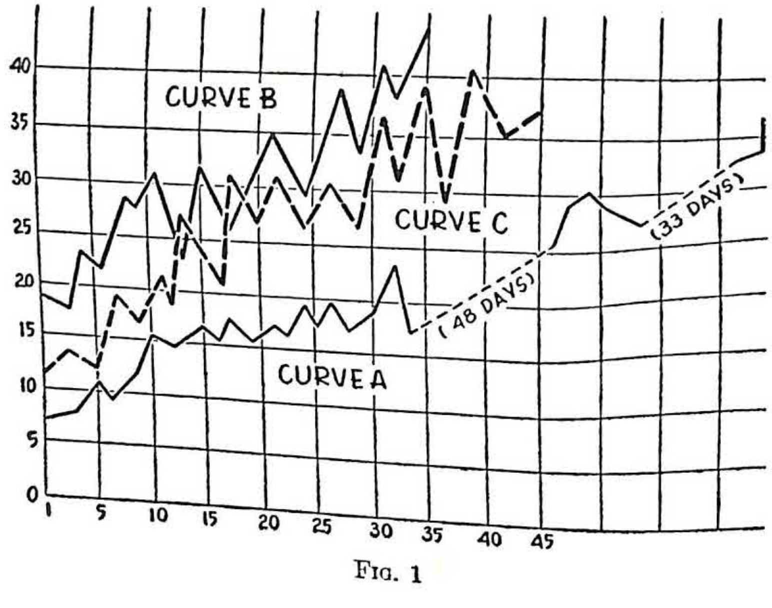

FIG. 1

words were arbitrarily deducted for each error.) In the first recall he wrote the same paragraph at the rate of 19.8 words per minute, and at the beginning of the second recall at the rate of 10.8 words per minute. These measures taken at face value indicate that the experimenter, on the first day of the first recall, wrote the paragraph at a rate that took twenty-seven days practice to attain in the original experiment. In the second recall, which was undertaken twenty-five years after the first with no practice whatever in the meantime, and fifty years after the first learning, it took approximately eight days to secure the same skill. However, the two curves are similar throughout. Near the ends of the two recall curves it may be seen that the experimenter, in twenty-five to thirty days, was performing in a manner equivalent to that which had taken one hundred twenty-six days to accomplish in the original learning. At the end, both curves indicate that there was no limit to improvement in sight, and regardless of the age of the learner, improvement might go on indefinitely.

In copying the three hundred-word page of new material (Fig. 2), the story is somewhat different. In the original learning the experimenter copied the page at the rate of 4.4 words per minute. In the first recall he copied at the rate of 12.4 words per minute at the first practice, and in the second recall, on the first day, he wrote the page at the rate of 6.1 words per minute. Glancing at the two sets of practice curves it would seem that there is considerable similarity at the beginning. The experimenter was copying at a rate which, in three to seven days, took four times that many days to accomplish at the original learning. The first recall continues to show considerable improvement over the first learning, but the early promise of the second recall soon fades out and the curve

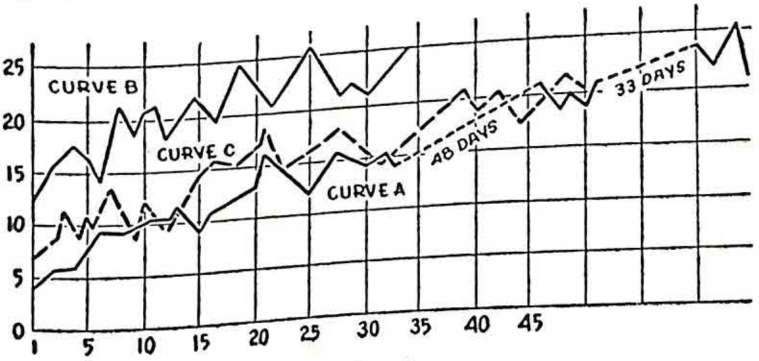

FIG. 2

almost parallels the original learning curve. Even this may be favorably regarded as significant considering the age of the experimenter. He believes, however, that the lack of more favorable progress is due to poor eyesight which makes reading from the printed page laborious and slow. This difficulty probably accounts for the leveling-off of the second recall curve, and probably indicates a limit to improvement in the rate of copying anything from the printed page where reading is involved.

Within the limits of this experiment certain observations may be made. The experimenter believes that age, as such, is not necessarily a deterrent to either new learning, or to recalling that which has formerly been learned. Physical handicaps do necessarily make a difference in performance. Muscular reflexes, where physical handicaps are not involved, perhaps may present a question for further investigation and study. This experimenter is inclined to believe that muscular reflexes in the aged, as in youth, show individual differences.

Some of the same factors that are present in learning at an earlier age are also present and effective at maturity. Interest, desire, and determination are potent factors. The will, as such, is a powerful stimulus. Competition against one's self is also very effective and may be more so in the aged than youth. While competition against rivals or competitors is a great stimulus to youth, the aged, as a rule, do not care who excels them.

In the case of the present experimenter, when he discovered that he could recall and even excel his previous record, and that he possessed the necessary skill to typewrite (for all practical purposes), he did not feel that extra exertion for more skill was worthwhile; although he does feel that he has by no means reached the limits of his capacity to improve in typewriting skill.

Referring again to the learning curves A, B, and C, we see ups and downs representing good and relatively poor performances, (as are found in all records of learning). The introspective notes which were kept during the experiment, along with casual observations, indicate that these variations in performances may be due to a current lack of enthusiasm for the task, a feeling of well-being—or the lack of it (sometimes referred to as the mood of the learner), or they may be due to some hidden factors in human performances which have not yet been discovered.

RECRUITING PROSPECTIVE RURAL TEACHERS AT THE MEN'S TEACHERS' COLLEGE TRIPOLI, LIBYA

IBRAHIM E. METAWEH[1]

UNESCO Mission, Tripoli

During the writer's successive contacts as a UNESCO technician with the Libyan education officials, the problem of effectively recruiting prospective rural teachers was raised. Teachers and students at the Teachers' College were included also in these contacts, and as a consequence, the idea of establishing an adequate instrument for recruitment was defined and clarified. The writer had noticed in the College several striking cases wherein students with serious defects had been admitted in the past, and since it seemed desirable to have some kind of device that would be effective in screening applicants, the idea of the interview test gradually took shape in the minds of the educational officials involved.

The purpose would be to raise the standards of prospective rural teachers admitted to the Teachers' College by means of a practical, objective instrument for screening applicants, namely, an interview test—one which might be standardized sufficiently to be used in other Teachers' Colleges in the country. The hoped-for result would be the selection of the fittest for the teaching profession from among the applicants, most of whom come from rural districts. The procedure adopted may be of interest to those working in similar regions and perhaps to others as well.

THE INTERVIEW CARD

Information, such as age, previous school grades, address, name of parents, etc., was obtained from every student applying for admission. Space was provided for each rater to mark the applicants on each of seventeen items. The items appearing on the card are presented below, supplemented by explanations and samples of the questions asked. The total number of points or marks arbitrarily assigned to each item and (for the main points) the number of questions under each are shown in parentheses. Students scoring between 40 and 49, were to be referred to a revision committee composed of the Dean of the College and two teachers in the Education Department. The minimum total to be obtained by every student was set at 50 marks out of a total of 100.

---

[1] Technician of Rural Education, UNESCO Mission, Tripoli.

(1) *Physical Aspects (30 Points, 24 Questions)*

Failure in this item alone might result in ruling out the student. This would happen if he has an apparent physical defect sufficient to hinder him from expressing himself or from performing activities necessary to the teaching profession, or to make him marked among his students. This applies to the cases where the college cannot do anything at present to correct the defects. In such cases, however, the interview is continued and the student is referred to the revision committee. Elimination of cripples is necessary as we need the teacher to be a leader too, but this does not rule out the possibility of his being accepted in another profession.

(a) *General appearance (10)*. Personal cleanliness and being well dressed without regard to the quality or price of clothing worn.

(b) *Speech and clarity (10)*. Notice in the students' answers the fluency of speech, diction, ease in pronunciation, and form of sentences.

(c) *No tics and defects (10)*. Notice here apparent lameness, dwarfishness, partial but apparent, or complete deafness, cross-eyedness, stuttering, stammering, lisping, awkward-walking, one-eyedness, facial deformities, and various tics, such as exaggerated gesturing, excessive movement of eye lashes, winking, excessive movement of head or shoulder and the scratching of various parts of the body.

(2) *Attitudes and Personality (25 Points, 28 Questions)*

(a) *Emotional balance and control (5)*. (The student can be asked questions like the following, although it is not necessary to confine the questions to those mentioned here.) (*a*) What is your reaction if someone in a group does not like you? (*b*) What is your reaction if somebody differs from your opinion, and what are your feelings toward him? (*c*) How would you behave if you knew that one of your friends told others something about you that you did not like? (*d*) If you were late to a lecture and the teacher refused to admit you, how would you behave? (*e*) What would you do if you were to fail this test?

(b) *Social interests (5)*. (*a*) Do you like to be alone most of the time or do you get bored? Do you like to join groups? (*b*) Do you like to participate with others in organizing parties and public meetings? (*c*) Do you prefer to have a picnic alone or with others? To what extent? (*d*) What is the number of your acquaintances and friends? (*e*) Do you like to break the rules?

(c) *Leadership and guidance (5)*. (*a*) If you are asked to participate in organizing a party, trip or social event, what rôle do you like to take? (*b*) If you like to accompany others on a picnic, do you like to have an opinion as to where you should go, or do you accept easily any place suggested by the others? (*c*) If you walk with others, do you like to be in front or in the middle, or do you not care about your order? (*d*) Have you studied the life of a leader? Who is he, and what did you admire about him? (*e*) If you are among a group of people who are discussing some affairs, what is your stand? Do you just listen to others? Do you participate in the discussion or do you guide the discussion?

(d) *Ability to comprehend and adjust (5)*. (*a*) Complete the following sentence: "The relationship between the ox and the plough is like the relationship between the locomotive and . . . ." (*b*) Complete the following chain of numbers:

9–1–9–2–9– . . . – . . . – . . . – . . .
2–6–10–14– . . . – . . . – . . . – . . .
3–7–15–31– . . . – . . . – . . . – . . .

(*c*) Ali is Ibrahim's cousin on both sides (mother and father). What is the relationship between the father of Ali and the mother of Ibrahim? (*d*) The speed of Ali is half the speed of Ahmed and Ahmed's speed is one and a half as much as the speed of Elsayed. What is Ali's speed if Elsayed's speed is 6 kilometres per hour? (*e*) What would you do if the neighbouring house were on fire? (*f*) What would you do if you were asked to talk on a subject with which you were not acquainted? (*g*) Supposing that you were late for an important appointment where your presence was necessary, how would you act? (*h*) If you were given fifty Libyan pounds, what would you do?

(e) *Independence and self-reliance (5)*. (*a*) Do you seek the help of others regarding your personal problems? (*b*) If you find it difficult to understand a lesson or an arithmetic problem, what figures about these states? The relationship between Libya and these states? Important products of these states? Their famous men? Kind of government? The United Nations, Specialized Agencies and Security Council? The functions of each?

PROCEDURE AND APPLICATION OF THE TEST

Each interviewee was rated by a committee of at least three persons from the staff of the Teachers' College, one with a scientific background, the second with an educational or social-studies background, and the third with a linguistic or artistic background.

Directions to the committee members caution them not to begin to question a student until they establish rapport and are sure that he is at ease. Students can be assured of permissiveness by talking to them about simple and ordinary things in a friendly way. It is strictly forbidden to scold a student or to be sarcastic to him during the interview, no matter what his answers are. He is not to be hurried in giving his answers or interrupted in his train of thought. The committee members are reminded that the students examined are young and in the process of growth, and that the conditions of their environment, their way of life, and economic status should be taken into consideration while examining them. It is not advisable to ask the students about peculiar features of areas that they have not visited. To give them adequate time to reveal their interests and attitudes, and at the same time to guard against exhausting them, the period of the interview was held to ten or fifteen minutes.

Members of the testing committees are free to start the test without adhering to the consecutive order of questions given. They might like to start by inquiring of the student about his locality. Also, it is not compulsory to follow the patterns of questions and directions stated. They are given as examples but the test permits flexibility.

When the test interview with its four main points, one hundred marks, and one hundred and forty questions was finally agreed upon, the writer presented a copy with the instructions to the appropriate authorities. A committee of four that was formed to examine it approved it without modifications. Fifty copies were mimeographed, twenty-seven of which were distributed among the teachers who conducted the test. Eight examining committees were formed at the Teachers' College made up of three teachers each, all of whom studied the test in advance.

RETURNS AND EVALUATIONS

Five hundred and thirty-six students applied for admission varying in age from eleven to twenty-four, approximately three out of every four students coming from rural areas. The average age of the students who applied was sixteen years and ten months. The average age of the admitted students was sixteen years and nine and a half months. Twenty-six students were not admitted because they were thought to be too young or too old. (Table I.) The age of seventeen years was thought to be most suitable for prospective teachers as they will have to stay in college for four years, and by the time they graduate at the age of twenty-one, it is highly probable that they will be mature enough for teaching.

Of the five hundred and ten applicants not disqualified because of age, 60.4 were admitted (57.5% of the total number). It may be that some of the more than one-third who were disqualified might have become successful in college and as teachers. This will never be known, and only time will tell how many of those who were admitted will be successful. Of those disqualified because of a score of less than 50, some extreme cases were—ten students who were extremely short, four one-eyed, seven stuttering and stammering

TABLE I.—AGE OF APPLICANTS ADMITTED AND NOT ADMITTED

Age	Applicants	Admitted	Not Admitted
12 years or less	10	0	10
13 years	9	5	4
14 years	9	5	4
15 years	34	21	13
16 years	149	91	58
17 years	181	108	73
18 years	93	58	35
19 years	25	13	12
20 years	6	4	2
21 years	4	3	1
22 years and over	16	0	16
Total	536	308	228

cases, five lame, and twenty apparently suffering from malnutrition.

This project is radical in recruiting intended teachers in this region where admission procedures have long remained unorganized. There were some doubts as to the necessity for this new interview test among some officials and teachers since they were well enough satisfied with conditions as they were. Throughout the test and after it was tried out their attitude changed. The writer is glad to report that they are now enthusiastic about it and are willing to participate in every phase of it in the coming years. This in a way proved the effectiveness of the test.

Since it had not been customary to obtain the exact birth dates of applicants nor a transcript of credits from the schools they had previously attended, it is as yet impossible to correlate ages and marks with the results from the test interview, but it is proposed to correct for this deficiency in the future. Similarly to test the instrument's validity correlations will be obtained with the results of the diploma examinations and with marks in practice teaching. Certain revisions will be made in the interview test to include more questions, particularly in the field of artistic interests, which were not covered. Work is soon to start on developing an intelligence test, the results of which will be used to supplement the other information used as criteria for admission.

Lack of social interest, independence, and self-reliance were apparent from the interview test results coupled with the observations of the testing committees. The reasons for this are difficult to ascertain, but it is felt that the lack of extra-curricular clubs might, in part, be responsible. So far there are five newly formed clubs in the College: the Historical Club, the Educational Club, the Journalism and Magazine Club, the Athletic Club, and the Scientific and Cinema organization. Although they are cultural in nature, social attitudes, independence, and self-reliance are involved.

The first use of an instrument in this country has been most gratifying since it was a coöperative enterprise that seemed to meet the needs of the situation as we knew it. As a result of its use and of other projects under way, we can look forward to a gradual upgrading of the teachers in Lybia, and hence, of its educational system.

THE INTELLIGENCE TEST IN SCHOOL USE: SOME PERSISTENT ISSUES[1]

J. B. STROUD

State University of Iowa

The first successful intelligence test was designed with little regard to definition or theory. Binet proceeded in an empirical fashion to ascertain kinds of tests that would discriminate among criterion groups. Kuhlmann, Terman and most of the early workers adopted the Binet procedure. On the other hand we are all familiar with various analytical approaches to the problem of intelligence test design. Thorndike's *area, altitude* and *speed*; Spearman's *g* and *s* factors; and Thurstone's *primary mental abilities* are cases in point. Thus, Spearman wrote, "When a knowledge of *g* and *s* factors makes it possible to construct tests which measure *g* and *s* at the maximum, an accurate measure for *G* for any individual can be obtained." This, he maintained, is far better than an average derived from a miscellany of unrelated traits (7).

Thus, the history of the intelligence test reflects two major approaches in test design: (1) the empirical and (2) the analytical. In the former, the psychologist attempts to maximize differences among criterion groups. This is largely a utilitarian approach. In the analytical approach the psychologist investigates the nature of intelligence, with little immediate concern for the practical. Obviously it was Spearman's hope that the two approaches would ultimately amount to the same thing. As yet this hope has not been fulfilled. Test designers have generally resorted to the Binet method of maximizing differences among criterion groups. The advancement in design is reflected largely in statistical and logical refinement of Binet's general plan. Perhaps the most notable departure is the PMA battery. Here one gains the impression, from the somewhat limited research, that insofar as the tests of the PMA battery are factorially pure, most of them assess mental abilities not closely correlated with school achievement (*11*).

Binet regarded the MA as a general average. Perhaps the general average is still the most significant feature of the major tests.

---

[1] Presidential Address, Division of Educational Psychology, 1956.

However, most general tests have featured batteries of subtests permitting part scores and resulting profiles. This practice tends to insure systematic sampling of intellectual abilities and thus to improve the "general average." Ostensibly the purpose of this feature is to provide for differential diagnosis and prediction. Although discussion of the analytical features of intelligence scores has sharpened within the last decade or so, the usefulness of the part scores in school situations still suffers from a lack of detailed research on the differential power of the constituent subtests.

Nonverbal tests came into the picture originally as a means of examining persons presenting language handicaps. Most of the new general intelligence tests within recent years have featured both verbal and nonverbal scores. Wechsler's recent and impressive Intelligence Scale for Children, affectionately referred to as the WISC, and the more recent and equally impressive Lorge-Thorndike tests are cases in point. Insofar as we know, at least insofar as I know, this trend in test design is dictated more by fashion than by empirical evidence attesting its usefulness in general school situations. I suspect one would have some difficulty marketing group intelligence tests today that did not contain nonverbal features. I also am inclined to suspect that their worth as nonverbal tests in school practice has not been demonstrated by research. What the clinical psychologist is able to make of the results of this kind may be another matter.

Most of the earlier nonverbal tests were seriously lacking in validity as intelligence tests. We have just completed a correlation analysis of the WISC on about eight hundred school children, grades three to six inclusive. In our results the nonverbal tests predicted academic achievement quite as well as the verbal tests. The attested usefulness in school situations, however, of the Wechsler nonverbal tests may well stem from the fact that they are valid tests of intelligence, rather than from their nonverbal character. Quite conceivably additional verbal tests of comparable quality might be substituted for the nonverbal ones without detriment to the battery.

Our analysis failed to yield much in the way of differential prediction of the achievement measures used, reading, arithmetic, and spelling. It is true that four or five of the eleven subtests gave better prediction than did the others of the three achievement measures. But this apparently resulted from the fact that they were better tests rather than from qualitative differences. A multiple regression analysis was performed based upon within-grades intercorrelations between the eleven WISC subtests and the three achievement measures. *Beta* weights were tested for statistical significance and least significant variables successively eliminated. It was found that four WISC subtests—Arithmetic and Vocabulary among the verbal tests, and Block Design and Object Assembly among nonverbal tests—accounted for most of the correlation between the WISC and the achievement measures. The within-grades least squares multiple correlation between all eleven WISC tests and reading, for example, was 0.74. The within-grades least squares multiple correlation between these four subtests, Vocabulary, Arithmetic, Block Design and Object Assembly, and reading was 0.73. The correlation between the unweighted raw score sums of the four subtests just referred to and Reading was 0.69. Comparable results were obtained for arithmetic achievement and for spelling achievement. It is a particularly striking fact that for all practical purposes the same four WISC subtests are the most effective predictors of each of the three achievement tests. For cross validation purposes the same kinds of analyses were carried out on a sample different from that on which the regression equations were derived, but drawn in the same way. The results obtained with the independent sample were strikingly similar to those obtained on the original sample (*12*).

Incidentally, analyses of this kind seem to be indicated for differential aptitude tests. It has been suggested that in view of the high correlations obtained between general aptitude tests and intelligence scores and also between general achievement tests and intelligence scores, the intelligence test is no longer required in school practice insofar as prediction of achievement is concerned. It is true that prior achievement predicts future achievement at least as well as does the general intelligence test. But it is equally true that the intelligence test which can be administered in a fraction of the time and at a fraction of the cost predicts future achievement almost as well. The merits of the argument must rest pretty largely upon the nature of the evidence of differential prediction. This brings us back to the intelligence score. How much have we improved upon Binet's "general average," beyond obtaining better averages? The value of pat scores on a test—verbal, nonverbal or others, depending upon the analytical properties of the test—must be assessed in terms of differential prediction, actually substantiated by research. Perhaps it is true that the general achievement test and the general intelligence test each tells us something unique about the pupil and that arguments about one's replacing the other could well be forgotten. However, and for the time being, it may be well to bear in mind that the intercorrelations among achievement scores in various school subjects are relatively high. In view of this fact it will be difficult to do a great deal with differential prediction. Consider a medical aptitude and a legal aptitude test, for example. One has a medical aura, the other a legal one. But it still may be true that they could be pretty well interchanged so far as their basic purposes are concerned.

Another issue at least related to the intelligence score is the generality of the mental age. The mental age score, whether determined by Stanford-Binet method or that derivable from Wechsler tests by multiplying the obtained IQ by CA, is, in effect, an average score. Are some MA's, say, of ten, better than others? Do MA's of a given magnitude vary qualitatively? In our work we have sought to determine whether or not MA's of a given magnitude vary qualitatively with the brightness of the pupils achieving them. A related but somewhat different question is whether or not we can make the same predictions relative to current school achievement of bright and dull pupils earning like MA's. Of course, we would predict much greater future achievement for the younger pupils. In one investigation Kolstoe (*8*) formed matched pairs of bright and dull pupils achieving equal, estimated true Stanford-Binet MA's. The mean Stanford-Binet MA was 11.3. The mean CA of the dull members was 15.5, giving a mean IQ of 79. The mean CA of the bright members was 9.3, giving a mean IQ of 126. The bright and dull pupils in the sample thus formed were then compared on the basis of the eleven WISC subtests and the PMA subtests. Except for PMA subtest Number, differences between the bright and dull pupils on the seventeen subtests were small and generally not significant. The dull, older pupils were far better in performance on PMA subtest Number.

In a companion investigation, Bliesmer (*2*) compared bright and dull pupils, matched in MA, on reading achievement. On reading subtests Recognition of Main Ideas, Drawing Inferences, and Total Comprehension, the bright, younger pupils excelled by significant amounts.

The school placement of the older pupils in this sample was three or more grades higher than that of the younger pupils. Owing to the hierarchical character of the arithmetic curriculum, the younger pupils had had no instruction in some of the operations which the older pupils had studied. This may account for some or all of the superiority of the older pupils in Number. One might think, *a priori,* that young, bright pupils would surpass (in achievement) older, dull pupils of comparable MA's. They are likely to be better motivated, they have been expected to do well in school, they are more likely to be accustomed to success. In general they have had better early home training and have had more encouragement and stimulation at home. Of course, these advantages may be offset by the fact that at the age level at which the work just mentioned was done the older pupils had spent twice as many years in school as the younger ones.

At the present time we are witnessing a growing interest in the education of gifted pupils. This is obliged to sharpen interest in the nature of the mental abilities of the gifted, especially the highly gifted. Perhaps we would be inclined to infer from the results of the work on the generality of the Mental Age that ten-year-old pupils earning MA's of fifteen would behave on such tests as the WISC and the PMA much as fifteen-year-old pupils of average intelligence.

The work on the mental age concept just referred to certainly suggests that for the age ranges involved, the MA as a unit of measurement possesses a high degree of generality. This is not to defend the Stanford-Binet procedure of designing intelligence tests. I do mean to defend the MA as a highly useful unit of measurement. The Wechsler method of designing intelligence tests, while it may not produce a superior result, appears to be more efficient and more economical.

The logic of the argument that the MA as a unit of measurement suffers from known or suspected inequalities of its units is not impressive. The meaning and usefulness of Height Age or Weight Age or Dental Age are unaffected by the fact that these units prove to be unequal when tested by a metrical scale. A height age of nine is scarcely more meaningful in school work because we know how much this is in metrical units. A Mental Age, or for that matter a Reading Age, of nine is not less meaningful because we are unable to say how much intelligence or how much reading ability it is.

Moreover, the inequality of successive MA units, or our lack of knowledge regarding their equality, argues nothing with respect to the equality of IQ's obtained at successive age levels. The IQ defines a pupil's mental development in relation to the mental development of other pupils of like age in the general population. In a sense it defines the rate of mental growth. We may say that a six-year-old pupil who achieves an IQ of 125 develops at the rate of 1.25 of a mental year for each calendar year. At this age his attainment equals that of the average pupil seven and a half years old. He has gained a year and a half in six years. He would be expected to gain another year by the time he is 10. We could also say he is approximately 1.5 S.D.'s above the mean for his age, or that he exceeds about 93 per cent of the pupils of his age in the general population. The fact that the MA score from which the IQ is derived at successive ages may not be equal has nothing to do with a pupil's maintaining his relative position in his age group.

Terman's concept of validity of the IQ implies an absence of systematic variations in IQ with age. Given constant means and S.D.'s over an age range, a given IQ must signify the same degree of brightness at successive age levels within this range. This assumes nothing with respect to the equality of the MA units from which the IQ's are derived. Unfortunately, test designers have not always been as careful as they should have been to publish S.D.'s for successive ages over the range of years covered. In the absence of this or comparable information we cannot tell whether or not an IQ of a given magnitude at one age level is strictly comparable numerically to one of like magnitude at another age level.

One of the more troublesome practical problems is the comparability of IQ's obtained on different tests—one might say needlessly troublesome. It should not have been a difficult matter at all for test designers to have provided school psychologists or other school officials with the variability data necessary to the establishing of numerically equivalent scores on various tests.

In school work there constantly arises the question of the significance of fluctuations in IQ from one testing to another, no matter whether the same or different tests are used on different occasions. When test designers, as Terman and Wechsler have done, provide the necessary data, the school psychologist can readily acquaint teachers with the amount of fluctuation required on a given test to become significant at given levels of confidence. There is no excuse for test authors' not doing this.

Traditionally our schools have been interested in several kinds of growth data. The variables for which we have the most satisfactory measures are mental growth, educational growth, and physical growth. It is of interest to note that these three types of growth possess considerable stability. Correlations between intelligence scores earned at eight and at eighteen years of age run around 0.85, as seen, for example, in Bayley's work (*1*). Correlations between measures of stature, obtained at eight and at eighteen years are of somewhat comparable magnitude (*9*), as are those between measures of educational achievement taken ten years apart.

We also know from an abundance of research that intelligence scores and educational achievement scores are rather closely intercorrelated, and that neither of these is closely intercorrelated with measures of physical growth. Moreover, we can improve prediction of educational achievement somewhat by combining intelligence and past achievement scores. We do not, as the work of Gates (*6*) showed more than twenty-five years ago, improve prediction by combining intelligence scores with various measures of physical growth.

However, the statistical method ordinarily used, that of multiple correlation, precludes our obtaining a lower correlation than that obtained between intelligence alone and achievement. If, on the other hand, we form simple averages involving MA's and various physical growth age scores, such as Height Age, Weight Age, Grip Age, and so on, and correlate the averages thus obtained with achievement, we do stand to obtain a much lower correlation. Indeed, if as many as five or six physical age measures are thus combined with MA and the resulting averages correlated with achievement, the coefficient should approach in magnitude the coefficient obtained between physical growth measures and achievement.

This is not to controvert the importance of physical growth in the conduct of schools. Physical growth features in the health and well-being of children. It is probably an important factor in their social and emotional adjustment. Certainly it would appear that any serious deviations should pose personal problems to which adjustment must be made. But we do not improve, but lower, prediction of achievement by combining physical age scores with MA scores.

For about as long as I have been acquainted with the subject at hand, there has been some expressed concern about the effect of reading proficiency on group intelligence scores. To be sure this is primarily a practical problem. But I suspect it could be argued that some of us have published papers on problems that had less practical—or theoretical—significance. In a way it is a bit strange that more work has not been done on this problem, if for no other reason than that of allaying the concern of teachers and school officials.

Some work of one of our own students on an MA thesis suggests that a slow reader who has good reading comprehension is penalized, so to speak, on timed group tests *(10)*. No evidence was obtained to suggest that the pupils scoring low on reading comprehension, whether fast or slow in rate, were penalized in any way. It seems probable that a facile reader who enjoys excellent reading comprehension would outperform on timed group intelligence tests, an equally intelligent pupil who reads slowly. Whether such an outcome would help or hinder prediction would depend upon the relationship between reading rate and educational achievement. One would suspect that the relationship is far from close. It is conceivable, as Durrell *(4)* pointed out long ago, that since reading proficiency is correlated with achievement, the presence of a reading factor in group intelligence tests may actually improve their predictive power.

Obviously, we would not expect the effect of reading proficiency on group intelligence test scores to be the same for all such tests. We would anticipate that the obtained effect would, among other things, prove to be a function of the speed-power variable.

By way of closing, some brief remarks are made apropos of the social status question in intelligence testing. Without doubt, the publicity given this question by the work of Eells, Davis and colleagues *(5)* and the earlier work of Cattell *(3)* should have a wholesome effect upon test construction. As a group, test authors in the past were probably guilty of some degree of carelessness with respect to this problem. It is perhaps too early to attempt to resolve the important issues raised. The need is urgent to understand the meaning of social class differences in test scores as psychological phenomena. The use of such expressions as "culture fair" and "culture bias" gives the impression that these differences are the unwitting work of test authors. I say unwitting because surely no one would have done this by design. Obviously, if these differences are chargeable to test authors, they are unimportant as psychological phenomena, save as they become something to be studiously avoided. On the other hand, we may find that the problem is really too big for the test author. It may turn out that the kinds of cultural impact associated with social class differences affect the course of *mental development* of children as well as their performance on intelligence tests.

Social class differences are real differences—substantial psychological phenomena with which schools and society must deal. It may be that we cannot build valid intelligence tests which will not at the same time discriminate among the social classes. Or, if we start the other way round by designing tests which will not discriminate among social classes, we may find that the tests are poor predictors of academic achievement.

REFERENCES

1) N. Bayley, "Some Increasing Parent-Child Similarities During the Growth of Children," *Journal of Educational Psychology,* 45, 1954.

2) E. P. Bliesmer, "A Comparison of Bright and Dull Children of Comparable Mental Ages with Respect to Various Reading Abilities," *Journal of Educational Psychology,* 45: 321–31, 1954.

3) Raymond B. Cattell, "Classical and Standard Score IQ Standardization of the I.P.A.I. Culture-Free Intelligence Scale 2," *Journal of Consulting Psychology,* 15: 154–59, 1951.

4) D. D. Durrell, "Influence of Reading on Intelligence Measures," *Journal of Educational Psychology,* 34: 165 ff., 1943.

5) Kenneth Eells, et al. *Intelligence and Cultural Differences, Chicago,* University of Chicago Press, 1951.

6) A. I. Gates, "The Nature and Educational Significance of Physical Status and of Mental, Physiological, Social and Emotional Maturity," *Journal of Educational Psychology,* 15: 329–58, 1924.

7) Katharine B. Greene, "General Intelligence and Its Measurement," *Review of Educational Research,* 2: 277, 1932.

8) O. P. Kolstoe, "A Comparison of Mental Abilities of Bright and Dull Children of Comparable Mental Ages," *Journal of Educational Psychology,* 45: 161–168, 1954.

9) H. V. Meredith, "The Prediction of Stature of North European Males through the Elementary School Years," *Human Biology,* 8: 279–83, 1936.

10) J. F. Rummel, "The Effect of Reading upon Group Intelligence Test Scores," Master's Thesis, University of Iowa, 1947.

11) D. C. Shaw, "A Study of the Relationships between Thurstone Primary Mental Abilities and High School Achievement," *Journal of Educational Psychology,* 40: 239–49, 1949.

12) J. B. Stroud, Paul Blommers, and Margaret Lauber, "Correlation Analysis of WISC and Achievement Tests," *Journal of Educational Psychology,* 48: 18–26, 1957.

THE INFLUENCE OF INTENTION ON THE RETENTION OF SCHOOL MATERIALS

DAVID P. AUSUBEL, SEYMOUR H. SCHPOONT,
and LILLIAN CUKIER

Bureau of Educational Research

University of Illinois

Extrapolating from general principles of educational psychology and from related experimental evidence, many educators and psychologists implicitly believe that intention to remember facilitates the retention of learned verbal materials. Actually all of the evidence bearing on this issue is ambiguous and inconclusive for one or both of two reasons: the experimental arrangements have been such as (1) to induce intentions to learn rather than to retain, or (2) to make impossible the isolation of the effects of intention on learning from its effects on retention. The present study was designed to overcome these interpretive ambiguities by both introducing *explicit* intentions to remember for an extended period of time and by experimentally isolating the effects of such intentions from original learning. The latter result could only be insured by introducing intentions to remember *after* completion of the learning task.

Numerous studies have shown that intention to remember influences the quality and longevity of retention. When subjects learn material with the expectation of recalling it for a designated period of time, recall is superior for the expected as against either a shorter or longer interval (*2, 3, 8*). Lester demonstrated that retention is enhanced by expectation of recall and by foreknowledge of the occurrence and possible effect of interpolated materials (*5*). Unfortunately, however, since differential intentions to remember were introduced at the time of original learning, these experiments did not isolate the effects of intention on the quality or quantity of what was learned in the first place from its effects on retention per se.

Another group of studies (e.g., *6, 7*) has shown that retention is superior and retroactive inhibition less marked when practice is accompanied by "intent to learn" than when learning takes place incidentally. That this difference is largely attributable to superior original learning, however, is demonstrated by the fact that it no longer prevails when Ss are equated for original mastery of the learning task (*1*). Heyer and O'Kelly (*4*), on the other hand, showed that learning under conditions of strong ego-involvement did significantly increase retention even though their ego-involved and control groups could not be differentiated on the basis of original learning outcomes. It is quite possible, however, that the superiority of the ego-involved group on original learning was masked by the inhibitory effect of intense ego-involvement on initial test performance. In any event, the evidence yielded by all of these studies can at best be indirect, because in each instance an explicit experimental set was induced to *learn* material for immediate reproduction rather than to *retain* it for an extended period of time. Furthermore, even if ego-involved intention to learn leads to no demonstrable quantitative superiority in original learning, it may still affect retention by influencing certain qualitative aspects of the learning process.

METHOD

Standardization of learning material and tests of retention. The learning material used in this study was a passage of approximately one thousand four hundred words on the history of opiate addiction. It was selected on the basis of being generally unfamiliar and inaccessible to undergraduate students, but yet learnable and relatively interesting. Two equivalent multiple choice tests on this material were constructed as follows: fifty-six five-choice test items were administered to a group of fifty-five undergraduate students in education three days after they had studied the passage for twenty-five minutes. Following item analysis, the order of items was rearranged so as to match odd and even items with respect to difficulty and type of question. The revised form of the test was then administered to a different group of twenty-nine undergraduate students in education. The range of scores (number correct) on this test was from thirteen to forty-five with a mean of thirty-three and a standard deviation of 7.8. The uncorrected odd-even coefficient of equivalence was 0.76. For the main experiment the odd and even items respectively constituted the two equivalent forms.

Experimental procedure. Our experimental population consisted of ninety-seven undergraduate students enrolled in four sections of an educational psychology course at the University of Illinois. Of these, fifty-three were in the control group and forty-four in the experimental group. In order to minimize the possibility and consequences of communication between groups, the control group was handled during the Fall semester and the experimental group during the following Spring semester. The experiment was rationalized as a laboratory exercise and was performed during scheduled class hours. The control Ss were first instructed to read the passage at their normal reading speed (approximately eight minutes), and then to utilize the remainder of the twenty-five available minutes to commit the facts and concepts in the passage to memory. It was explained that a multiple choice examination on the material would be given immediately after the reading was completed. Ss were assured that the results on the examination would not be made public and would not affect their grades in the course. They were promised, however, that their scores would be returned to them individually, and that the median and frequency distribution of the class scores would be placed on the blackboard so that they could compare their own scores with the class norms. Fourteen days after the original learning and test, an unannounced (surprise) retest was given, utilizing the equivalent form of the test. Hence the control group had no explicit intention to remember the material for the second test.

During the second semester the same procedure, with one exception, was followed for the experimental group. Immediately *after* the Ss finished the first test, they were informed that a second test on the same material would be given in fourteen days. This announcement presumably induced explicit intentions to retain the previously learned material for an additional two weeks without in any way affecting the quality or quantity of original learning.

RESULTS

Means and standard deviations of learning (test) and retention (retest) scores for control and experimental groups are shown in Table I. Comparison of test and retest scores indicates that a significant degree of forgetting ($p < 0.001$) occurred in each group during the two week interval. The mean learning and retention scores made by the two groups were also significantly greater in all

TABLE I.—COMPARISON OF CONTROL AND EXPERIMENTAL GROUPS ON LEARNING AND RETENTION TESTS

	Control Group		Experimental Group	
	Test	Retest	Test	Retest
Mean	20.1	13.2	19.5	13.6
S.D.	2.53	3.06	2.56	2.58
Corrected mean*	13.4	6.5	12.8	6.9
Percentage retained†	48.5		53.9	

* The corrected means were computed by subtracting 6.7 (the mean score of subjects who took the test without studying the material) from the group means of raw scores (number of correctly answered items).

† Corrected means were utilized in computing the percentage of retention from test to retest.

instances than the test scores made by different Ss who answered the test questions without prior study of the learning passage. The product-moment correlation between test and retest scores was 0.45 ($p < 0.01$).

Since the means, standard deviations, and variances of the two groups are not significantly different from each other on the learning test, we are justified in comparing control and experimental groups with respect to percentage of retention from test to retest. The respective percentages of 48.5 and 53.9 show only a negligible difference in degree of retention which does not even begin to approach statistical significance.

DISCUSSION

The results of the present study do not support the widely held view that intention to remember facilitates the retention of learned verbal material. In reconciling these findings with previously reported studies in this area (*2, 3, 5, 8*), we can only surmise that intention to remember in these latter studies did not influence the retention process per se, but affected various quantitative and qualitative aspects of original learning which indirectly increased the availability of the learned material at a later date.

Obviously, if a student learns material more thoroughly in the first place because he expects to be tested on it several weeks later he will have more available to remember than a student who originally learns less diligently because he does not expect to be retested. However, if the intention to remember is introduced *after* learning takes place it apparently has no influence on the longevity of the learned material. Intention to remember, in other words, affects retention by enhancing learning rather than by increasing the stability of existing memory traces; or, to put it negatively, intention *not* to remember material beyond the date of the examination leads to little subsequent retention not because of any direct influence of this intention on memory traces but because such intentions are invariably associated with superficial and cursory methods of original learning. Hence, teachers who wish to increase the longevity of learning by maximizing intention to remember can only do so by inducing the appropriate intention at the time of original learning. Admonitions to remember once learning is completed, even if they lead to corresponding intentions in the student, do not alter the subsequent fate of memory traces.

One possible qualification of these findings should be considered. Intention to remember in our Ss might have been relatively weak in the absence of tangible material rewards for superior retention. High retention scores, after all, did not "count" in the sense of improving students' grades. On the other hand, level of ego-involvement was maintained at a reasonably high level by intrinsic desire to coöperate in a research study relevant to the subject matter of the course and by Ss' incentive of scoring favorably in relation to class norms.

SUMMARY AND CONCLUSIONS

An attempt was made to ascertain whether explicit intention to remember facilitates retention of learned school materials apart from its influence on original learning. An experimental group of undergraduates studied a historical selection for twenty-five minutes and was tested on same immediately afterwards by means of a multiple choice test. *After* this test, an explicit intention to remember was induced by announcing that an equivalent form of the test would be given two weeks later. The same procedure was followed with a control group except that an unannounced retest was given.

The two groups were not significantly different in mean learning scores or in the percentage of material retained from test to retest. It was concluded that intention to remember facilitates retention by enhancing original learning rather than by increasing the stability of existing memory traces.

REFERENCES

1) W. C. Biel, and R. C. Force, "Retention of Nonsense Syllables in Intentional and Incidental Learning," *Journal of Experimental Psychology*, 32: 52–63, 1943.

2) F. D. Boswell, and W. S. Foster, "On Memorizing with the Intention Permanently to Retain," *American Journal of Psychology*, 27: 420–426, 1916.

3) M. T. Geyer, "Influence of Changing the Expected Time of Recall," *Journal of Experimental Psychology*, 13: 290–292, 1930.

4) A. W. Heyer, Jr. and L. I. O'Kelly, "Studies in Motivation and Retention. II. Retention of Nonsense Syllables Learned Under Different Degrees of Motivation," *Journal of Psychology*, 27: 143–152, 1949.

5) O. P. Lester, "Mental Set in Relation to Retroactive Inhibition," *Journal of Experimental Psychology*, 15: 681–699, 1932.

6) J. Peterson, "The Effect of Attitude on Immediate and Delayed Recall: a Class Experiment," *Journal of Educational Psychology*, 7: 523–532, 1916.

7) W. C. H. Prentice, "Retroactive Inhibition and the Motivation of Learning," *American Journal of Psychology*, 56: 283–292, 1943.

8) M. N. Thisted, and H. H. Remmers, "The Effect of Temporal Set on Learning," *Journal of Applied Psychology*, 16: 257–268, 1932.

THE FACTORIAL INVARIANCE OF PURE-FACTOR TESTS[1]

JOHN W. FRENCH

Educational Testing Service

Current practice in assembling batteries of pure-factor tests has been to select, for each factor, one or two tests that have been found by factor analysis to have high loadings on that factor and low loadings on all others. Factorial invariance is assumed. That is, if several tests were to be selected to represent each of a number of factors and tried out with a new group of subjects, it would usually be assumed that the factors would reappear clearly defined by the expected tests. The analysis described in this paper was carried out to check on the invariance of a set of pure-factor tests selected for possible use in making educational predictions.

The tests making up the experimental battery of pure-factor tests were suggested by tests used in the studies covered in a survey of the factor-analysis literature (*2*). In general, three tests were selected as the purest and most reliable measures of each of thirteen of the best-established aptitude and general achievement factors.

Deviating from this pattern, Reading Comprehension and Mathematics Reasoning were used as measures of two factors each: Verbal Comprehension and General Reasoning for the former, Number and General Reasoning for the latter. Although these two are not pure-factor tests, they were included because of their known validity for a number of purposes. The Carefulness factor represents another deviation from the pattern of three tests for each factor. Since not much confidence was felt in the way that the scores for the Carefulness factor would turn out, five tests were tried. Four are used in the analysis; the other was eliminated, since eighty-five per cent of the subjects received a zero score. It should be pointed out that the Carefulness scores were derived from the tests of other factors. However, it was believed that there would be little distortion of the factors produced by experimental

---

[1] Grateful acknowledgment is made to the staff of the U. S. Military Academy for their coöperation in scheduling and administering the tests and in supplying data on course grades. Without this kind of coöperation the study would not have been possible.

dependence, because the four Carefulness scores bore no consistent relationship to the scores on the four tests from which they were derived.

THE TESTS

The tests are described in the order in which they were administered. The first four were a part of the entrance tests given five months earlier than the experimental battery. The expected factor content is given, as are the number of items, the time limit, and an estimate of speededness. The word "speeded" is used to indicate that very few or no subjects finished. "Some speeding" indicates that it was estimated that twenty-five per cent to seventy-five per cent finished. Most of the tests in the experimental battery were given in two separately-timed parts in order to make possible the computation of alternate-form reliability coefficients. The coefficients corrected by the Spearman-Brown formula are given with the test descriptions. A full description of the test is not given when the name of the test or the kind of test is familiar to most readers (*2*).

For Verbal Comprehension

Verbal

(1) *Antonyms (Entrance)*. 40 items, 15 minutes. 93% finished.

(2) *Reading Comprehension (Entrance)*. 20 items, 25 minutes. 93% finished. This test was intended to measure both Verbal Comprehension and General Reasoning.

(3) *Verbal Relations (Entrance)*. Verbal-analogies items and verbal-completion items. 40 items, 30 minutes. 88% finished.

For General Reasoning

Reading Comprehension. (Described above.)

(4) *Mathematics Reasoning (Entrance)*. 20 items, 30 minutes. 61% finished. This test was intended to measure both General Reasoning and Number.

Variables 5 to 44 were scores from the experimental battery.

(5) *Figure Analogies.* Two parts, each 20 items, 10 minutes. Some speeding. $R = 0.64$.

For Space

(6) *Similar Figures.* From four squares containing black and white areas the one was to be selected which was not simply

rotated but was the mirror image of the other three. Two parts, each 40 items, three minutes. Speeded. $R = 0.89$.

(7) *Block Counting.* In pictures of large piles of similarly shaped blocks, the subjects were to count the shaded blocks that touched each numbered, unshaded block. The contacts between the blocks had to be inferred from the parts of the blocks shown. Two parts, each of five piles and 50 items, five minutes. Some speeding. $R = 0.82$.

(8) *Revised Minnesota Paper Form Board.* Series MA (Psychological Corporation). One of five patterns was to be selected which could be produced by putting together given outlined parts. 60 items, 25 minutes. Speeded.

For Speed of Judgment

(9) *Visual Preferences.* Each item consisted of a pair of somewhat similar, simple line drawings or figures. The subject checked the one he liked best. The score was the total number checked. No account was taken of the kinds of figures preferred. (For this test and the following two tests, approximately every fifth item presented a very easy choice so that subjects responding at random could be detected. No such subjects were found.) Two parts, each 50 items, two minutes. Speeded. $R = 0.88$.

(10) *Social Judgment.* Each item consisted of pairs of names of personal qualities. The subject checked the quality he would prefer to characterize someone with whom he had to associate closely. The score was the number checked. No account was taken of the kinds of qualities preferred. Two parts, each 70 items, two minutes. Speeded. $R = 0.92$.

(11) *Size Judgment.* Each item consisted of pairs of descriptions of objects. The subject checked the larger (assuming standard or average sizes). The score was the number checked. No account was taken of the correctness of the judgments. Two parts, each 50 items, two minutes. Speeded. $R = 0.92$.

For Perceptual Speed

(12) *Picture Discrimination.* Each item consisted of three simple drawings of faces, two exactly alike, and one different in some respect. The subject checked the different one in each item. Two parts, each 56 items, two minutes. Speeded. $R = 0.89$.

(13) *Cancellation.* Each part consisted of a page of typed, ran-

dom, capital letters reproduced in red. The task was to draw an X over every A, starting at the top of the page and working down. The score was the number of A's cancelled. Two parts, each two minutes. Speeded. $R = 0.96$.

(14) *Number Checking.* This test is similar to the Minnesota Clerical Aptitude Test. Two parts, each 48 items, two minutes. Speeded. $R = 0.89$.

For Number

Mathematics Reasoning (Described above.)

(15) *Checking Addition and Division.* This was an alternation of five items, each of simple addition and division of 1- and 2-digit numbers, with an answer given. The answer was to be checked as correct or wrong. Two parts, each 50 items, two minutes. Speeded. $R = 0.89$.

(16) *Subtraction and Multiplication.* This was simple subtraction and multiplication of 1- and 2-digit numbers in spiral, omnibus form with five response alternatives. Two parts, each 42 items, two minutes. Speeded. $R = 0.91$.

For Carefulness (with these variables the signs of the correlations were reversed, so that high scores would represent favorable performance)

(17) *Picture Discrimination (Wrong).* This was the wrongs score on Variable 12. $R = 0.56$.

(18) *Checking Addition and Division (Wrong).* This was the wrongs score on Variable 15. $\mathrm{R} = 0.42$.

(19) *Subtraction and Multiplication (Wrong).* This was the wrongs score on Variable 16. $R = 0.38$.

(20) *Cancellation (Omit).* This was the number of uncancelled A's prior to the last A cancelled in Variable 13.

For Visualization

(21) *Similar Rotations.* The same solid form is shown in two different rotational positions in each item, labeled 1 and 2 respectively. A new form, labeled 3, was to be visualized as going through the same rotation in one or two planes as that represented by drawings 1 and 2. The resulting position of form 3 was to be selected from drawings A, B, C, D, or E. Two parts, each 20 items, five minutes. Some speeding: 53% finished Part 1; 41% finished Part 2. $R = 0.69$.

(22) *Folded Paper Test.* This test was similar to Thurstone's

Paper Folding Test (*2*). Two parts, each 20 items, five minutes. Some speeding. $R = 0.71$.

(23) *Mechanical Movements.* Drawings were presented of different kinds of mechanical apparatus involving gears, belts, eccentrics, etc. One or two items based on each drawing asked about the direction, speed, or nature of the movement within the apparatus pictured. Two parts, each 10 items, eight minutes. Some speeding: 25% finished Part 1; 31% finished Part 2. $R = 0.68$.

For Induction

(24) *Inducing Rules.* This test presented, in spiral omnibus form, items requiring the subject to discover a rule and then to answer the item by applying it. Included were analogies, series, and belonging items drawn from verbal, social studies, science, numerical, and graphic material. Variety was introduced to prevent the subjects from using a system for solving the problems. Two parts, each 24 items, seven minutes. $R = 0.72$.

(25) *Letter Sets.* The set of four letters not belonging with other sets of four letters was to be chosen. Two parts, each 15 items, five minutes. Considerable speeding: 7% finished Part 1; 20% finished Part 2. $R = 0.43$.

(26) *Locations.* This test was adapted from Thurstone's Marks Test. Each item presented five groups of dashes with spaces between them. An X replaced one dash in each of the first four lines, following some rule for placement. The subject was to indicate where the X should appear in line 5. Two parts, each 24 items, five minutes. Speeded. $R = 0.72$.

For Fluency of Expression

(27) *Metaphors.* This was similar to Thurstone's Similes. The subject was required to complete a simile in three different ways. The score was the number of completions regardless of their quality; no entirely senseless ones were made. Two parts, each 10 items, three minutes. $R = 0.84$.

(28) *Double Opposites.* This was similar to Thurstone's test of the same name. Each item presented a word and called for writing two words opposite to it in meaning and beginning with specified letters. The score was the total number of correct responses. Two parts, each 25 items, four minutes. $R = 0.72$.

Ednl.
Dated

(29) *Multiple Expressions*. Part I required the subject to write down as many ways as he could of requesting a second helping of meat. Part II required the subject to write down as many ways as he could of asking someone for his name. In each case the score was the number of expressions written down regardless of their quality; no senseless ones were made. Two parts, each four minutes. $R = 0.82$.

For Aiming or Eye-hand Coördination

(30) *Dotting*. The subject placed a dot inside as many typewriter-sized capital O's as he could. The O's were irregularly spaced along crooked lines. The score was the number of O's correctly marked. Very few were incorrectly marked. Two parts, one minute each. $R = 0.92$.

(31) *Tracing-Easy*. A line was to be drawn as rapidly as possible through lines ¼ inch apart having 3⁄16 -inch openings. The score was the number of lines correctly passed. Very few were passed incorrectly. Two parts, one minute each. $R = 0.92$.

(32) *Tracing-Difficult*. A line was to be drawn as rapidly as possible through lines 1⁄10 inch apart having ⅛-inch openings. The score was the number of lines correctly passed. Very few were passed incorrectly. Two parts, one minute each. $R = 0.94$.

For Motor Speed

(33) *Writing X's*. This score was the number of X's drawn on or near regularly spaced dots. Two parts, one minute each. $R = 0.92$.

(34) *Writing "lack."* This score was the number of times the word "lack" could be written (or printed) on or near regularly spaced lines. Two parts, one minute each. $R = 0.92$.

(35) *Writing Digits*. This score was the number of times the digits 1 to 9 could be written on or near given lines. Two parts, one minute each. $R = 0.91$.

For Mechanical Knowledge

(36) *Mechanical Information*. This was a multiple-choice factual information test in practical mechanics. Two parts, each 12 items, five minutes. Power conditions. $R = 0.74$.

(37) *Tool Information*. Each item presented the picture of an object used in some mechanical operation and the pictures

of three tools or other objects. The task was to select the one of the three tools which was commonly used with the first object. Two parts, each 15 items, five minutes. Power conditions. $R = 0.80$.

(38) *Mechanical Comprehension.* This was patterned after the Bennett test. Two parts, each 15 items, five minutes. Power conditions. $R = 0.64$.

Variables 39–44 were scores taken from a ninety-item multiple-choice information test designed to measure interest in six areas on the assumption that a subject who has been sufficiently interested in an area to participate in it or read much about it will have picked up factual information about it. The scores are the number of items correct among the 15 (placed in random order) representing each area. Total time, 50 minutes. Unspeeded. The areas covered were:

(39) *Interest Information (Science).* Common-sense science.
(40) *Interest Information (Social).* Games, dancing.
(41) *Interest Information (Financial).* Economics, banking, stock-market.
(42) *Interest Information (Athletic).* Baseball, sailing, bowling, etc.
(43) *Interest Information (Art and Music).* Painting, great music, technical information.
(44) *Interest Information (Mechanics).* Home shop tools, auto mechanics.

THE SUBJECTS AND THE ADMINISTRATION

The battery, five and one-half hours long including directions and administrative time, was given to about seven hundred and sixty plebes shortly before the end of their first summer at the United States Military Academy at West Point. With a ninety-minute history examination and a twenty-minute biographical inventory, it was given in two half-day sessions on successive days. In order to include some of the entrance test variables in the study, it was necessary to restrict the group to those who had been "Regular" and "Validating" candidates. This left out the "Certificate" candidates, i.e., those Congressional appointees who are certified by their schools. The number of subjects with scores on all forty-four variables was four hundred and ten.

THE CRITERION VARIABLES

Five additional variables as follows were obtained when first-year course grades became available in the summer of 1951:

(45) *Language Grades.* Choice within quotas of French, German, Portuguese, Russian, or Spanish. Fundamentals of grammar, reading, writing, and particularly oral practice.

(46) *English Grades.* Includes review of fundamentals, composition, reading of literature with class discussion and themes, and speech making.

(47) *Military Topography and Graphics Grades.* Use of instruments, engineering drawing, freehand and design sketching, map projections, descriptive geometry, and charts and graphs.

(48) *Mathematics Grades.* Includes algebra, trigonometry, solid geometry, analytic geometry, and an introduction to calculus.

(49) *Aptitude for Service Ratings.* Ratings on leadership potential. These are made by an officer advisor and by all fellow cadets in the cadet's company including those from all four classes. The officer's rating counts one-third; the cadets' ratings count two-thirds.

The product-moment intercorrelations were computed for the 44 variables on the four hundred and ten cases. For the tests which were divided into two parallel parts, the score used for the intercorrelations was the sum of the part scores.

For all correlations with the criterion variables the number of subjects was three hundred and sixty-one. This reduction was caused by normal attrition that took place during the first year at the Academy. The intercorrelations and the means and standard deviations may be obtained from ADI (*8*). Attrition at the Academy during the first year tended to raise the means and decrease the standard deviations a little, as might be expected. For tests given in two parts the alternate-form reliability coefficients corrected by the Spearman-Brown formula are given in the diagonal of the table.

THE FACTOR ANALYSIS

Factor analysis of the 49-variable intercorrelation matrix was carried out by means of Thurstone's Grouping Method (*5*). The highest value in each column was used as an estimate of the communality.[2]

Sixteen factors were extracted before the residuals seemed to be reduced to near the chance level. The groups of variables used for the extraction of the sixteen reference factors may be obtained from ADI (*8*).

After the sixteen factors had been extracted, no pattern that seemed significant could be found among the residuals. The largest value among the residuals was 0.079. The distribution of the residuals and the unrotated factor matrix may also be obtained from ADI (*8*).

THE ROTATION

Rotation to oblique simple structure was carried out by the method described by Thurstone (*5*, p. 194), where all pairs of factors are plotted on paper and the new positions for the hyperplanes are drawn on the paper, so that the direction cosines for the new rotation may be read from the graph. Seven sets of rotations were carried out, including a total of one hundred and twenty rotated pairs. A new set of graphs was assembled after each set of rotations. The final transformation matrix may be obtained from ADI (*8*). The final oblique solution is presented in Table I. Loadings of 0.20 or higher are printed in parenthesis. Communalities derived from the unrotated matrix are given in the right-hand column. Table II shows the intercorrelations of the factors.

In the process of rotation it was borne in mind that the infinite number of rotational solutions simply represent alternate ways of explaining the intercorrelations that were found. No one solution is more mathematically correct than another. While simple structure was the general goal, this concept is so variously or vaguely defined that there is much freedom for variation within its limits. Rotation in this study was not done blind, but was carried out so that the result would be of maximum value to the broad purpose

---

[2] After the extraction of the ninth factor, the values used for the communalities were the highest values in the columns of the residual table. Up to the extraction of the ninth factor the highest values in the columns of the original table were inserted in the diagonals of the residual tables. Since no groups had overlapped up to this point, it made little difference whether the communalities of members of the groups were estimated from the original or the residual tables.

TABLE I—ROTATED FACTOR MATRIX AND COMMUNALITIES
(Decimal points omitted. Loadings of 0.20 or higher are printed in parenthesis.)

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	h²
1 Antonyms (entrance)	(66)	−03	−04	−02	09	07	−04	−10	08	−01	−01	−02	−02	−03	04	19	79
2 Reading comprehension (entrance)	(43)	05	06	−01	−09	−07	10	06	−12	−02	−03	−01	04	08	−05	(−23)	68
3 Verbal relations (entrance)	(55)	−03	−03	04	00	01	−07	04	04	03	05	03	−02	−05	02	07	76
4 Mathematical reasoning (entrance)	04	04	−08	−03	−03	01	05	02	05	−10	09	(32)	−04	02	00	07	54
5 Figure analogies	02	04	−01	−04	09	06	−05	14	15	01	−04	18	06	12	−07	−01	52
6 Similar figures	01	−05	−01	06	−09	−03	−02	(23)	00	03	00	−06	00	(30)	04	06	45
7 Block counting	−10	−08	11	01	−01	−00	−00	(28)	−11	−07	08	03	11	(22)	−03	07	54
8 Minnesota paper form board	−10	04	18	04	−01	−03	14	(32)	00	−01	−00	−12	−03	(24)	−05	−17	57
9 Visual preferences	−12	−07	−01	−02	(59)	04	05	03	06	04	−10	−01	−01	04	01	−06	66
10 Social judgment	−01	−02	−03	02	(65)	−01	−00	−02	02	−04	−00	−07	04	02	−00	05	75
11 Size judgment	04	07	01	−01	(48)	−03	−05	−02	−05	00	18	07	−03	−05	−01	00	72
12 Picture discrimination	06	19	01	02	19	02	−05	13	05	05	−01	−07	−03	16	−01	04	65
13 Cancellation	03	(36)	02	00	−02	−02	04	−03	−18	−06	05	05	01	01	−05	05	49
14 Number checking	−03	(32)	−03	−02	03	01	−03	05	−01	04	−01	−04	−06	03	04	−07	64
15 Checking addition and division	02	−01	−01	−02	01	03	01	00	−06	01	−01	−01	−02	−01	(43)	04	85
16 Subtraction and multiplication	−00	02	−02	03	−02	−04	−02	−01	03	−01	−02	−00	02	02	(42)	−05	84
17 Picture discrimination (wrong)*	−04	02	04	−03	−05	−12	01	05	(37)	07	01	−03	−01	01	−04	(26)	41
18 Checking addition and division (wrong)*	03	05	02	−04	−01	03	03	−08	(46)	−01	−05	06	−01	−04	−04	−04	37
19 Subtraction and multiplication (wrong)*	−03	−03	−04	−03	05	07	−03	−04	(43)	−05	05	01	−08	04	03	(−22)	37
20 Cancellation (omit)*	−08	−09	01	09	01	−06	04	−01	(31)	−03	01	−00	12	−01	−01	07	23
21 Similar rotations	−06	08	02	−06	10	−01	02	−04	15	−05	05	−01	02	(31)	01	−04	40
22 Folded paper test	−04	01	07	−07	02	10	03	06	16	01	−06	11	−05	(27)	−01	−04	53
23 Mechanical movements	02	08	11	−00	03	−02	−03	−09	−01	−09	10	13	01	(28)	−03	13	60

24 Inducing rules	05	−02	−07	−03	05	03	−04	(30)	03	11	(20)	06	−03	01	−06	06	64
25 Letter sets	−03	08	06	−07	04	−02	07	(29)	−02	13	−06	08	09	−07	−03	−03	40
26 Locations	03	05	−08	−00	06	07	−03	(30)	−04	−01	−06	(21)	−06	01	−06	03	49
27 Metaphors	−11	−06	02	02	−02	00	−02	−02	−05	(56)	04	−00	01	−00	05	02	54
28 Double opposites	(25)	07	−01	−01	06	−01	−02	−09	06	(28)	01	−02	02	00	03	03	36
29 Multiple expressions	08	03	−01	−00	02	02	05	04	01	(51)	−07	03	−04	−00	−05	−05	47
30 Dotting	−06	−04	−00	09	01	(41)	02	02	03	02	03	−15	05	10	04	−01	58
31 Tracing-easy	01	07	−08	02	−02	(56)	−08	01	−04	−03	05	14	−10	−09	−06	08	70
32 Tracing-difficult	−05	−05	08	−11	−02	(60)	06	−03	02	01	−08	01	05	−02	01	−06	61
33 Writing X's	13	02	04	(47)	−04	05	−03	−03	−07	−04	−01	−01	00	−02	04	18	51
34 Writing "lack"	01	04	−02	(60)	05	01	03	−07	08	02	−01	04	−00	−01	−07	−16	68
35 Writing digits	−06	−07	−01	(68)	−01	−06	01	10	−01	03	01	−03	01	03	02	−05	71
36 Mechanical information	06	−02	(44)	−02	−01	00	−01	−09	11	−04	(29)	−02	−05	−01	03	02	73
37 Tool information	−01	−01	(62)	01	05	−05	10	02	−01	−03	02	01	11	03	−01	01	75
38 Mechanical comprehension	06	−11	(39)	−03	−00	07	−01	06	07	08	−03	−09	02	(23)	10	−03	57
39 Interest information (science)	−04	03	02	−04	−05	02	00	−05	−01	−02	(48)	07	−04	−02	−09	01	50
40 Interest information (social)	06	−02	03	−01	01	02	−04	−07	−01	05	−04	−03	(48)	10	02	05	49
41 Interest information (financial)	08	−14	09	02	02	−07	−03	−01	−01	05	15	01	05	09	10	00	29
42 Interest information (athletic)	−06	01	−05	05	04	−03	06	09	−03	−02	04	08	(29)	−11	−07	−04	27
43 Interest information (art and music)	14	01	02	−05	−00	04	−07	04	09	−03	(25)	−18	11	03	01	−04	37
44 Interest information (mechanics)	−09	01	(43)	05	−02	−05	−02	01	02	03	(30)	05	01	−08	−09	01	63
45 Language grades	05	−04	−08	01	−02	01	(51)	04	−02	−03	−01	−03	03	−02	01	06	60
46 English grades	(28)	12	−03	−04	06	−02	(37)	02	04	06	−01	18	−04	−15	−10	02	69
47 Military topography and graphics grades	−12	−02	(24)	05	01	04	(25)	12	01	04	06	11	−05	(22)	−03	−07	74
48 Mathematics grades	−06	−01	02	01	−01	02	(45)	05	−04	−03	−03	(30)	07	04	−02	−05	71
49 Aptitude for service ratings	−06	05	(26)	−04	−00	−06	(38)	−06	−09	06	−17	04	(27)	−01	−02	−02	24

---

* For these variables the signs of the correlations have been reversed, so that these variables represent Carefulness.

TABLE II—INTERCORRELATIONS OF THE FACTORS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1		−.02	.05	.01	.01	.05	.27	.35	.13	.24	.60	.46	.31	.39	.04	−.20
2	−.02		.03	.42	.38	.55	.26	.58	.16	.18	.27	.11	.33	.23	.74	.05
3	.05	.03		.05	.17	.18	−.23	.06	−.10	−.03	.24	.19	−.22	.37	−.11	.14
4	.01	.42	.05		.35	.54	.18	.13	−.07	.30	.18	.07	.11	.06	.36	−.01
5	.01	.38	.17	.35		.34	.13	.25	−.32	.49	.27	.07	.20	.08	.41	.03
6	.05	.55	.18	.54	.34		.18	.36	−.11	.19	.24	.07	.24	.27	.34	.07
7	.27	.26	−.23	.18	.13	.18		.23	.09	.15	.33	.08	−.02	.09	.34	.14
8	.35	.58	.06	.13	.25	.36	.23		.37	.16	.47	.46	.43	.48	.48	.01
9	.13	.16	−.10	−.07	−.32	−.11	.09	.37		−.28	.12	.25	.26	.15	.18	−.04
10	.24	.18	−.03	.30	.49	.19	.15	.16	−.28		.31	.13	.26	.07	.24	−.09
11	.60	.27	.24	.18	.27	.24	.33	.47	.12	.31		.51	.38	.43	.41	−.15
12	.46	.11	.19	.07	.07	.07	.08	.46	.25	.13	.51		.30	.58	.36	−.20
13	.31	.33	−.22	.11	.20	.24	−.02	.43	.26	.26	.38	.30		.08	.41	−.18
14	.39	.23	.37	.06	.08	.27	.09	.48	.15	.07	.43	.58	.08		.08	.16
15	.04	.74	−.11	.36	.41	.34	.34	.48	.18	.24	.41	.36	.41	.08		−.11
16	−.20	.05	.14	−.01	.03	.07	.14	.01	−.04	−.09	−.15	−.20	−.18	.16	−.11	

of the research, that of identifying a useful set of pure-factor tests by proving that an expected factor structure would appear. The criterion of simple structure and a positive manifold of loadings was responsible for a great majority of the decisions made in the rotation, but two examples may be mentioned to show the way in which other conditions played a part.

(1) This study does not reveal Space, Induction, and Visualization as three separate factors. Rotation could have been such that any two of these three factors would appear as primaries with the third being a combination of the other two. The final rotation was made to show the spatial tests as loading both Induction and Visualization, because this situation had the readiest psychological interpretation and, therefore, served as the best framework for the development and understanding of pure-factor tests.

(2) The second group of variables used in the extraction of factors consisted of the three "perceptual speed" tests and the two "number" tests. The first two rotations left the fifteenth reference factor as a residual, while the second was distinctly divided into two clusters, the perceptual and the number tests. The hyperplanes were then rotated so as to pass through these two clusters. Thus, Factor 2 became perceptual only, and Factor 15, which was a residual factor, became the Number factor. Although the two new factors correlate 0.74, the clear division between them provides convincing evidence that factorial invariance has occurred. These two

factors can, therefore, be relied upon to appear and may well be useful dimensions in a testing program.

INTERPRETATION OF THE FACTORS

The factors are named to indicate their identification with factors found by other investigators. Most of them are now included in the *Kit of Selected Tests for Reference Aptitude and Achievement Factors* (*4*), which was developed by committees of investigators for the purpose of defining some of the better established factors. The selection of some tests for the *Kit* was based partly on the findings of this analysis.

Following each factor name there appears the "Universal Index Number" of the factor from a numbering system for factors which was initiated by Cattell (*1*).

Because this study was undertaken to confirm the establishment of familiar factors, which have been described elsewhere (*2* and *3* and original sources referred to therein), the factor descriptions and the discussions are brief.

Factor I is *Verbal Comprehension*, Universal Index No. T13: Knowledge of the English language, including the skill necessary both to read and to write it.

Factor II is *Speed of Symbol Discrimination*, Universal Index No. T12: The ability to find a known symbol amidst a mass of distracting stimuli. Tests 12, 13, and 14 were included to define the "perceptual speed" factor. However, "perceptual speed," thought by the writer at the time of testing to be a single entity, actually consists of at least two factors, one of which is speed of symbol discrimination. Another "perceptual speed" factor calls for speed in perceiving detail. Variable 12 is a test of this other factor. Its low loading seems to confirm that fact.

Factor III is *Mechanical Knowledge*, Universal Index No. T8: A knowledge stemming from instruction or experience in the field of mechanics. This factor is unique in this battery of so-called aptitude factors, since it depends on variance in achievement. It was included in this aptitude battery because mechanical knowledge plays such an important part in our culture that it is likely to be of consequence in predicting success in various academic or vocational fields.

Factor IV is *Motor Speed*, Universal Index No. T9: Speed in moving the fingers in complex patterns that are well fixed by habit. This may be the same as the U. S. Employment Service's Finger Dexterity factor (*2*), and seems certainly to be the same as Bechtoldt's "Motor Speed" factor (*2*), and Woodrow's "Speed" factor (*7*). It is not known to what extent this factor may be generalized beyond tasks carried out with the hand holding a pencil.

Factor V is *Speed of Judgment* (not indexed): Speed or lack of compulsiveness in making minor decisions. The large individual differences on tests of this sort are probably affected rather strongly by the situation and the test directions. However, it seems reasonable to assume that a part of this variance represents a mode of response habitual to the individual.

Factor VI is *Aiming* or eye-hand coördination, Universal Index No. T1: The three variables having high loadings on Factor VI in this study are all performed with pencil and paper. These tests occur in several studies in the literature without evidence that the factor encompasses more than the paper and pencil situation (*2*).

Factor VII is the *Grades* factor (not indexed): The variance that is common to course grades but not to any tests. This variance probably covers a wide variety of characteristics including such things as classroom attitude and parental pressure.

Factor VIII is a combination factor of *Induction* and *Space*, Universal Index Nos. T5 and T11: *Induction* involves the findings of general concepts that fit sets of data, the forming and trying out of hypotheses. *Space* is the ability to comprehend the nature of the arrangement of elements within a visual stimulus pattern primarily with reference to the location of the examinee. The zero hyperplane of this factor could have been rotated to a different position with respect to Factor XIV, but the present rotation is the one which is most meaningful. The *Space* factor appears between *Induction* and *Visualization* and lies in the plane determined by these two. This makes good psychological sense, as suggested by the term "Spatial Reasoning" so often used instead of "Space." The three variables (6, 7, and 8) put into this battery to represent the "Space" factor are shown here to be a combination of a reasoning factor and the factor called *Visualization*.

Factor IX is *Carefulness* (not indexed): The ability or disposition to proceed cautiously on speed tests. In all of these tests the nature of the task itself is easy, but the directions call for work at top speed. The table of intercorrelations indicates that there was no consistent correlation between the carefulness scores and the speed scores derived from the same tests. Because this is true, there can be no serious distortion of the factor pattern resulting from use in the analysis of two scores derived from each of four tests.

Factor X is *Fluency of Expression* or possibly *Idea Fluency.* (The latter would be Universal Index No. T6) : The three tests were placed in the battery to define *Fluency of Expression,* the ability, once an idea is given, to express that idea rapidly in many different ways. However, the distinction between new ideas and new expressions of an idea is not clear and is not established here, because no recognized test of *Idea Fluency* was included in the battery. Such a test, for example, might call for writing down as many ideas as possible about some such topic as "a house." The score on this kind of "Idea Fluency" test would be the number of ideas written down within a given short time limit.

Factor XI appears to be *Interest in Science* (not indexed) or an information acquisition factor. Even if it is considered an interest factor, it is not known whether it, defined as it is by tests of *information* about science, is the same as that found in numerous other studies loaded by questions asking about *preference* for science. The Inducing Rules test includes questions involving a knowledge of general science, and so its loading is reasonable, but there is no explanation for the loading of Interest in Art.

Factor XII is probably *General Reasoning,* Universal Index No. T34: A factor which has been associated with arithmetic reasoning tests and a rather varied list of other reasoning measurements. Figure Analogies was included in the battery to help define this factor. Its loading of 0.18 is lower than expected, but it is the highest loading for Figure Analogies on any of the factors.

Factor XIII is an interest or personality factor. It could be called *Interest in People* or it may be similar to the *Sociability* factor found in personality inventory analyses. Factors having both of these interpretations have been found in many studies (*3*).

Factor XIV is a combination of *Visualization* and *Space,* Universal Index Nos. T14 and T11: As noted for Factor VIII, the inference here is that the tests used for the "Space" factor actually require the abilities represented by both *Visualization* and *Induction. Visualization* may be defined as the ability to comprehend objects in space so as to manipulate them in the imagination.

Factor XV is the *Number* factor, Universal Index No. T10: Speed in carrying out simple numerical computations involving the four arithmetical processes. Mathematics reasoning tests ordinarily have a loading on this factor as well as on *General Reasoning.* In this analysis the mathematics variables do not load on *Number,* because *Number* and *General Reasoning* were rotated to a correlation of 0.36 with each other.

Factor XVI has no apparent psychological meaning. It can best be considered a residual factor.

CONCLUSION

Tests of each of thirteen well-established factors and some tests of general information were administered to an entering class at West Point. Together with scores from the entrance examinations and first-year grades, these were factor-analyzed to confirm the identification of useful pure-factor tests by finding factorial invariance when such tests are factored.

All thirteen of the sets of tests used in the battery for the designation of reference factors did in fact turn out to fall into clusters within the factor space. Twelve out of thirteen of the clusters defined factors as expected. Those three variables intended to represent *Space* did not load a separate factor for themselves but did form a thirteenth cluster located in the plane defined by *Induction* and *Visualization.* The *General Reasoning* factor is included among the twelve that were separated successfully. However, the Mathematics Test (No. 4) and Figure Analogies (No. 5) do not define the factor well by themselves. The factor does become adequately clear, however, because of the presence of Mathematics Grades (No. 48).

The other three of the fifteen non-residual factors make good psychological sense: (1) The *Grades* factor has often appeared in studies where several course grades and tests were included. (2) The *Interest in Science* factor is one that has been found several times in the literature (*3*). Leadership ratings and amount of information on social activities combine to define a factor. The implication here is that the leadership ratings are heavily weighted with sociability.

From the standpoint of the study's purpose, it may be concluded that factorial invariance held up, and, therefore, the selection of tests to represent certain factors was successful. The clusters are sufficiently tight to make it reasonable to represent factors in a battery by one test from each set. Thus, the study served as a pretest yielding factorial, reliability, and level-of-difficulty data upon which to base the selection of tests.

REFERENCES

1) R. B. Cattell, A universal index for psychological factors. Advance Publication No. 3, Laboratory of Personality Assessments and Group Behavior, Univ. of Illinois, Urbana, Ill., 1953.

2) J. W. French, "The Description of Aptitude and Achievement Tests in Terms of Rotated Factors," *Psychol. Monogr.*, No. 5, 1951.

3) J. W. French, *The Description of Personality Measurements in Terms of Rotated Factors*. Princeton, Educational Testing Service, 1953.

4) J. W. French, (Ed.), *Kit of Selected Tests for Reference Aptitude and Achievement Factors*. Princeton, Educational Testing Service, 1954.

5) L. L. Thurstone, *Multiple-Factor Analysis*. Chicago, Univ. of Chicago Press, 1947.

6) Ledyard R. Tucker, "A Semi-Analytical Method of Factorial Rotation to Simple Structure," *Psychometrika*, 9: 43–68, 1944.

7) H. Woodrow, "The Relation Between Abilities and Improvement with Practice," *J. educ. Psychol.*, 29: 215–230, 1938.

8) American Documentation Institute. A more detailed form of this paper (or extended version, or material supplementary to this article) has been deposited as Document number 5163 with the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. A copy may be secured by citing the Document number and by remitting $1.25 for photoprints, or $1.25 for 35 mm. microfilm. Advance payment is required. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress.

RETENTION OF GENERAL PSYCHOLOGY

W. J. McKEACHIE and DANIEL SOLOMON

University of Michigan

Whenever a group of teachers gather to commiserate about the futility of trying to educate the average college student, someone is sure to say, "Within a week after the final examination the students will have forgotten three-quarters of what they've learned anyway!"

Such statements, based on Ebbinghaus's classic studies of memory have seldom been tested in terms of the amount of gain in knowledge during a course retained over a period of time following the course.[1] Ebbinghaus, himself, discovered that meaningfulness made for improved retention. We might, then, expect retention to vary somewhat from course to course as well as individual to individual.

If an instructor is teaching an advanced course, how much should he expect students to remember from previous courses? This study had the practical objective of determining how much of students' learning in a course in general psychology was retained over a summer or more.

Procedure

At the University of Michigan a course in general psychology is a prerequisite for the course in educational psychology. In the fall of 1954 most of the students in the educational psychology course had taken the general psychology course either a year earlier or the preceding spring. Accordingly, a forty-item multiple-choice test was constructed from items on the general psychology final examinations for those two semesters. This test was given to students in the educational psychology course during the first

---

[1] Most studies of retention of learning in school have simply compared scores on an examination at the end of a course with scores on the same test at some later date. Such a procedure ignores the possibilities that students might score as well on many or all of the test items at the beginning of the course as they do at the end of the course. The model which seems appropriate is illustrated by the study of Frutchey (*1*) who found in a study of high school chemistry that general principles are retained better than specific terminology.

TABLE I.—PERCENTAGE OF CORRECT ANSWERS TO EXAMINATION ITEMS

	Pre-test	End of General Psychology Course	Beginning of Educational Psychology Course Sept., 1954	Per cent Retention
19 items common to June, 1954, and Educational Psychology tests	54.03	73.5	70.7	85.6
18 items common to Feb., 1954, and Educational Psychology tests	41.12	78.0	71.1	81.3

week of the fall semester.[2] Thirty-five of the students entering the educational psychology course had also taken the final the previous June. There were nineteen items common to the two tests. Eighty-one students had taken the final the previous February. There were eighteen questions common to these two tests.

In June the mean score on the nineteen common items was 14.0. In September the same students averaged 13.4. While test papers were not available for the February examination, item analysis data gathered for all students in the course indicated that the mean on the eighteen common items had been 14.0, while the mean score of eighty-one of these students in educational psychology eight months later was 12.8. Since students entering education are on the average no better than the average student in our course, this too indicates a high degree of retention.

This was a much higher degree of retention than we had anticipated. We began to wonder whether or not our results could be explained by the hypothesis that our test scores were not really measuring course achievement. Perhaps the students would have done as well on these items even without taking the course. To check this we gave the same test to a group of students at the beginning of the course. The results showed their scores to be significantly lower than those of the educational psychology students.

As Table I indicates, if we take the gain in test score from beginning to the end of the course as measure of learning, retention over a four to 8-month period is more than eighty per cent. While the possibility remains that the test measures some general gain in sophistication independent of the course, it seems fairly safe to conclude that material learned in the first course in psychology is retained for an appreciable period after the final examination.

---

[2] The authors appreciate the help of Dr. Robert Dixon in arranging for administration of this test. Charles Smith and Diedrich Snoek assisted in test scoring and analysis.

REFERENCE

1) F. P. Frutchey, "Retention in High-School Chemistry," *Journal of Higher Education,* 8: 217–218, 1937.

ASSESSMENT OF INTELLIGENCE LEVEL BY THE RORSCHACH TEST AS COMPARED WITH OBJECTIVE TESTS

BOLESLAW A. WYSOCKI

Marquette University[1]

Rorschach wrote that his test "... presents an intelligence test almost completely independent of previous knowledge, memory, practice and degree of education" (*5*, p. 183). This high estimate of blots as to assessing intellectual abilities is shared, almost unopposed, by most of the Rorschach experts. They contend that intelligence is one of the most important factors that can be measured with a great accuracy by the Rorschach test. Thus the purpose of this paper will be to find out to what extent the group method of administration proves or disproves Rorschach's well-known contention concerning the assessment of intelligence level by his ink blots.

TESTS

The following tests were used in the present study: (1) Group Rorschach, scored after Klopfer (*2*, *3*, *4*); (2) Verbal Intelligence Test, "S.P. Test 15," used by the British Army and consisting of one hundred equally weighted items; (3) Nonverbal Intelligence test, Raven's Matrices, consisting of sixty equally weighted items.

SUBJECTS

In order to make comparisons with Rorschach categories, the verbal and nonverbal intelligence tests were given to one hundred and thirty-two men and eighty-five women. The ages ranged from eighteen to forty-six with an average of 27.9 years for men, and from sixteen to forty-three with an average of 21.5 years for women. The average educational background was grade 10.3 and 9.6 for men and women respectively.

---

[1] Now at the Alliance College, Cambridge Springs, Pa.

TABLE I.—CORRELATION COEFFICIENTS BETWEEN RORSCHACH CATEGORIES AND INTELLIGENCE TESTS

Rorschach Categories	Verbal Intelligence Test		Nonverbal Intelligence Test	
	Men	Women	Men	Women
No. of R	0.440	0.389	0.389	0.450
W%	0.239	0.331	0.329	0.112
M%	0.272	0.368	0.266	0.389
F + %	0.356	0.111	0.129	0.111
F − %	−0.115	−0.174	−0.083	−0.053
A%	−0.291	−0.205	−0.124	−0.184

RESULTS

In order to establish validity, if any, of the group Rorschach method in assessing intelligence level, the several Rorschach categories were correlated with the two intelligence tests.[2] In all cases coefficients were calculated by the full product-moment formula (Pearson) and are shown in Table I.

Correlations between the several Rorschach categories and the intelligence tests of 0.228 and 0.283 are significant at 0.01 level for men and women respectively; whereas correlations of 0.174 and 0.217 are significant at 0.05 level for men and women respectively.

As Table I shows the highest positive correlation with intelligence tests, significant at 0.01 level (both the verbal and nonverbal) is N of R, that is, total number of responses for men and women. The second category showing positive correlation at 0.01 level of significance with both tests is W, a complete whole response. Next to the W is the M category (human movement), showing positive correlations at 0.01 level of significance with both tests for men and women. Good form F + %, gives only one positive coefficient significant at 0.01 level with the verbal intelligence test for the male group. The poor form, F − %, gives negative correlations not significant at 0.01 or 0.05 levels. Animal responses, A, did correlate significantly with the intelligence tests at 0.01 level for men and 0.05 level for women.

Next, the possibility of combining the scores in the various Rorschach categories was taken into consideration to see if a

---

[2] Coefficient of correlation between the verbal and nonverbal intelligence tests was 0.707.

TABLE II.—INTERCORRELATIONS OF INTELLIGENCE TESTS AND RORSCHACH CATEGORIES

	Verbal Int. Test	No. of R	W%	M%	F+%	F−%	A%
N. verbal Int. Test	0.707	0.424	0.273	0.203	0.193	−0.052	−0.191
Verbal Int. Test	—	0.409	0.290	0.179	0.209	−0.146	−0.229
No. of R		—	0.384	−0.083	−0.054	−0.067	−0.222
W%			—	.092	0.064	0.093	−0.196
M%				—	0.334	−0.140	−0.209
F + %					—	−0.503	0.087
F − %						—	0.044
A%							—

combination would be better than the best single score. For this purpose the Wherry-Doolittle test selection method was adopted (*1*, p. 435) to satisfy need. Thus, the matrix of intercorrelations between the two intelligence tests and six Rorschach possible indicators of intelligence was computed. The correlations are simple product-moment coefficients for the whole group of subjects, that is, for one hundred and thirty-two men and eighty-five women tested in the study. The correlation matrix is given in Table II. Correlations of 0.181 are significant at 0.01 level and correlations of 0.138 are significant at 0.05 level.

Next, corrected multiple correlations were calculated as follows:

Rv.1	=0.409	Rn.1	=0.424 (R of 0.181 is significant at 0.01 level)
Rv.14	=0.464	Rn.13	=0.492 (R of 0.212 is significant at 0.01 level)
Rv.146	=0.487	Rn.134	=0.499 (R of 0.234 is significant at 0.01 level)
Rv.1463	=0.487	Rn.1345	=0.503 (R of 0.253 is significant at 0.01 level)
Rv.14632	=0.490	Rn.13456	=0.505 (R of 0.269 is significant at 0.01 level)
Rv.146325	=0.487	Rn.134562	=0.505 (R of 0.283 is significant at 0.01 level)

In the above-listed multiple correlations 1 stands for No. of R; 2 for W %; 3 for M %; 4 for F+ %; 5 for F− %; and 6 for A %. The symbol v represents the verbal intelligence test, and n the nonverbal intelligence test.

om the review of these multiple correlations one may observe that by adding several Rorschach categories to the first one selected (with the highest correlations with intelligence tests) substantial gains are obtained increasing multiple R from 0.424 to 0.505 and from 0.409 to 0.490 for intelligence tests, both nonverbal and verbal, respectively. On the basis of the above multiple correlations it may be noted that, when taken together, four Rorschach categories, namely number of responses, human movement, good form and animal responses, constitute the best combination of intelligence predictors, as measured by intelligence tests.

SUMMARY AND CONCLUSIONS

An attempt was made to determine the validity of the group Rorschach in assessing the level of intelligence. In the first stage of this investigation, the group Rorschach results were compared with two intelligence tests, verbal and nonverbal, for men and women separately. Significant correlations, ranging from 0.272 to 0.450, were found for the following Rorschach categories: total number of responses, the percentage of whole responses, the percentage of human movement responses, good form per cent, and the percentage of animal responses. The correlations between the poor form level and intelligence tests were found not significant when calculated for men and women separately.

Further, a consideration was given to the possibility of exploring combinations of several scores instead of single ones. Thus, after computing the correlation matrix for men and women together, several corrected multiple correlations were calculated. It appeared that by successive additions of single Rorschach scores the multiple correlation was significantly increased up to 0.081 R points. It also turned out that the following Rorschach categories constituted the best combination of scores for predicting level of intelligence: number of responses, human movement, good form, and animal responses.

These results confirm those of Wishner (*6*) with regard to the individual Rorschach.

REFERENCES

1) H. E. Garrett, *Statistics in Psychology and Education*. New York: Longmans, Green and Co., 1947.

2) M. R. Harrower-Erickson and M. E. Steiner, *Large Scale Rorschach Techniques*. Springfield, Illinois: Charles C. Thomas, 1945.

3) B. Klopfer, et al. *Developments in the Rorschach Technique.* New York, Yonkers-on-Hudson, 1954.

4) B. Klopfer, *The Rorschach Technique.* New York, World Book Co., 1942.

5) H. Rorschach, *Psychodiagnostics* (translated from German by P. Lemkau and B. Kronenberg). Bern, H. Huber, 1942.

6) J. Wishner, "Rorschach Intellectual Indicators in Neurotics," *American Journal of Orthopsychiatry*, 18: 265–279, 1948.

BOOK REVIEWS

SAMUEL J. BECK. *The Six Schizophrenias.* New York: American Ortho-Psychiatric Association, 1954, pp. 237, plus viii.

An important contribution to Rorschach researches is Dr. Beck's volume on the six schizophrenias. The study purports to be, and is, a coördinated inquiry into the problem of schizophrenia as carried on by the three scientific disciplines—psychology, social work and psychiatry. This is the orthopsychiatric pattern for the multi-discipline approach in the study of human behavior. The researches, one with children and one with adults, were conducted in 1947–52 in the Michael Reese Hospital in Chicago. The aim of the research group has been to describe the schizophrenic person by operational trait data which are clinically and psychodynamically meaningful and to inspect these descriptions by sound method.

Six patterns of dynamic organization of the traits emerged and the volume is chiefly concerned with a description of details of the study and the results found.

Dr. Ray R. Grinker contributes the initial chapter called "Clinical Introduction" and in it he describes the step-by-step build up of the research and the clinical significance of it for research and therapy. Considered in terms of methods and approaches, the research includes: (1) The analysis of a sizeable series of normal adults and children—sixty children, fifty adults; (2) an analysis of the clinical histories and the Rorschach tests of non-schizophrenics; (3) an analysis of clinical history and study of clinical formulations with Rorschach correlates.

The Stephenson Q technique used for sorting of clinical data and the corresponding Rorschach data on each case is considered in a special chapter included concerning Q methodology by Prof. Stephenson.

It is Dr. Grinker's opinion that the researchers give us not only a better idea of dynamic process of schizophrenia, a better way of understanding the process of disintegration and attempts at reintegration, a better predictive of diagnostic view of schizophrenia, but also a much more satisfactory way of viewing the personality. Dr. Beck perceives the researches as having four implications: (1) Describes the schizophrenia as certain over-all dynamic adjustment patterns; (2) emphasizes latency in schizophrenia that is not necessarily psychosis; (3) bimodality; (4) postulate of least resistance. And he is happy about the fact that the test as a tool has withstood the reactions of the tough minded group of psychiatrists that helped in his research.

There is no questioning the value of these researches for stimulating further research, as well as adding to our knowledge concerning the Rorschach as a tool in the studying of personality. What can be questioned is the practicality of the approach in terms of making the most useful contribution towards developing insights for understanding and treatment of schizophrenia.

However, the frame of reference for the researches is understandable enough considering Beck's persistent interest in the Rorschach test.

H. MELTZER
Human Relations Research Foundation, St. Louis, Missouri

DELOS D. WICKENS and DONALD R. MEYER. *Psychology*. New York: Dryden Press, 1955, pp. 541 plus vi. $5.25.

IRWIN A. BERG. *Student's Workbook*. New York: Dryden Press, 1955, pp. 181 plus v.

HARRY P. BABRICK. *Teacher's Manual*. New York: Dryden Press, 1955, pp. 70.

Another text in Psychology is this book by Wickens and Meyer. In it the authors endeavor to present the essential facts about behavior as they are now known and also a systematic method of explaining which they consider the most fruitful, rigorous and promising. As far as intellectual orientation is concerned, the chief emphasis in this book is on the importance of learning and the use of learning concepts in the analysis of complex forms of behavior. The early chapters of the book deal with learning concepts and are captioned "Learning Fixation" and "Utilization of Learning." These are followed with chapters dealing with basic concepts in areas of motivation and perception, social behavior and personality development, dealt with in the light of the basic learning concepts as presented in the earlier chapters. The last few chapters deal with physiological foundations of behavior and a very brief last chapter represents an attempt to sum up. Each chapter is followed by a list of references. The book contains an index. The publishers have done an excellent job in printing and preparing a book of attractive appearance.

This new psychology text includes the usual chapters, though somewhat rearranged. Diagrams and pictures indicate a fair amount of resourcefulness used in putting the volume together to make ideas considered clear. References are relatively of a more recent origin. Behavior is described by the authors as "responses made to stimuli acting on the individual at the particular moment. Because of the individual's biological structure and the nature of his past experience, these stimuli will evoke certain psychological mechanisms—such as habits and motives, to mention only two—and certain kinds of responses will follow." This approach, the authors say, does not deny conscious experience, it simply states it will not use conscious experience as the main source of explanation. In short, what they are saying is that they are dominantly behavioristic.

Academic newcomers in the field are likely to be less critical of the general content and emphasis. People who have grown with the field are likely to miss certain experiments, illustrations and conclusions that they are used to seeing and thinking of as being fundamental in presentation of the subject of psychology. People who are psychoanalytically oriented will not find the content or emphasis in this book satisfying. Their treatment of "learning to learn," therapists and clinicians will find lacking in an emphasis on need for unlearning. In the authors' consideration of forgetting, they cover qualitative experiments in forgetting. Qualitative in the sense of learning and forgetting more meaningful life experiences is not mentioned by them. Homeostasis is used in a completely uncritical manner. Persons who think that it is well for students to know something about historical foundations of the field, including some of the founders, will find the book disappointing in this respect. Thorndike's name, for example, is not mentioned and some of the other leaders in the field are barely mentioned. Though the writing is relatively well done, the emphasis on experimental data gives certain chapters popular interest, a feeling of being dried up.

What the authors have tried to do in this book, to a fair extent they have achieved in doing. A fair amount of material usually included in a first course is treated here compactly and in a relatively simple manner. Chapters on physiological foundations at the end of the book are relatively well done, but the kind of material students expect to find in psychology books is treated in less humanized fashion than needs be. All in all they have done well what they consciously strived to do. And omissions and shifts of interest obviously are intentional. It is a safe book to use for a psychology text rather than the most advisable one.

Student's Workbook for use with Psychology by Wickens and Meyer is prepared by Irwin A. Berg. It includes a listing of concepts, terms and objective test exercises of each chapter. It also contains some demonstrations, exercises and formal statements for demonstrations and projects.

Teacher's Manual to accompany the Wickens and Meyer book is prepared by Harry P. Babrick of Ohio Wesleyan University. It includes a list of text questions and discussion questions for use by teachers. It also contains an annotated bibliography of recent literature relevant to the teaching of general psychology.

H. MELTZER

Human Relations Research Foundation, St. Louis, Missouri

MILTON GREENBLATT, RICHARD H. YORK, and ESTHER LUCILE BROWN. *From Custodial to Therapeutic Patient Care in Mental Hospitals.* New York: Russell Sage Foundation, 1955, $5.00, pp. 497.

"From Custodial to Therapeutic Patient Care in Mental Hospitals" is a publication of the Russell Sage Foundation, which was established in 1907 "for the improvement of social and living conditions in the United States of America." The psychiatrist on the team, Dr. Milton Greenblatt, writes the chapters on "Towards Therapeutic Community" and with the social worker, E. Brown, writes the conclusions. The psychologist, Richard York, writes Part II, which is concerned with experiments in ward patient care in Veterans Administration Hospitals, and Part III where the problem of motivating patients on chronic wards at Metropolitan State Hospital is considered. The book also contains an excellent bibliography on social treatment compiled by Dr. Frederick L. Wells. It also contains an index.

That mental hospitals are behind times is no news for people in the field and for an increasing number of other people. The hospitals whose function it is to administer to the sick minds, rather than to the sick bodies, of nearly three quarters of a million patients, continue to struggle with buildings of which some are fire traps and almost all are overcrowded. They struggle also with an inadequate staff and inadequate facilities of all kinds. Care is still largely in the hands of attendants of whom there are supposedly from 80,000 to 90,000 and many of these positions are held by men and women who have neither the needed aptitude nor experience. The purpose of this volume is not to tell a weeping story of neglect for mental hospitals, but to present a description of work done and work that can be done to reduce the present confusion and discouragement. It is basically an exploratory review in social treatment. The attitude of the authors is eclectic in nature. As they see it, the function of a mental hospital is to utilize every form of treatment available for restoring patients to health. And at worst, to aid them to live as nearly normal lives as possible within an institutional setting. The goal of the authors is to use both the personnel and physical resources of the hospital to aid in the development of competence in the patients. The description of ward and ward treatment included will be of interest to all people working with institutional cases or in institutions. It will be of particular interest and use to administrators of mental hospitals. In their assessments of hospitals, collaboration with community sources and resources, the authors show a nice balance between the sense of limitations and the continued planning for improvement. In their emphasis on the need for improvements of conditions in chronic wards, the authors are calling attention to a phase that at the present time still is perhaps too often neglected. The authors are perfectly willing to consider that another kind of social environment completely outside the hospital could be devised which is better suited to the needs of some chronic patients. Their hope is that from studies (of the kind reported in this volume) of work done on the wards might evolve decisions concerning greater differentiation of care and the opportunity, as least for the mental hospital to become a hospital in fact, as well as in name.

One thing is certain—in all the mental hospitals which have been modernized, particularly by the addition of a better trained staff, an encouraging number of patients with acute conditions have been discharged. And thanks to help from the Federal Government, as well as increased awareness on the part of many state governments, sights are being raised and at least the beginnings of promise are in view. The question facing us now, as described by the authors, is "whether hospitals will strive to maximize the use of such physical and social resources as exist, including resources in patients, to achieve this end; and whether the American public will become a more active partner in this venture by greatly increasing contributions and sustaining interest and service."

The description of the material here, as well as the attitude expressed should be of help to all people connected with or working with mental patients.

H. MELTZER

Human Relations Research Foundation, St. Louis, Missouri

C. WINFIELD SCOTT and CLYDE M. HILL. *Public Education Under Criticism.* New York: Prentice-Hall, 1954, $4.75, pp. 414 plus xiv.

HENRY EHLERS. *Crucial Issues in Education.* New York: Henry Holt & Co., 1955, $2.25, pp. 277.

To the wave of criticism of our public schools and education in general, there is only one answer—more and better education. And that includes education of educators and interested laymen to an awareness of the meaning and significance of the issues. By so doing, the editors of the two anthologies on education hope to contribute toward the rescue of America from the retreat from learning advocated by critics. The two volumes being reviewed are both concerned with the problem of how best to handle the criticism of education and how to develop realistic attitudes for action to improve the educational scene.

"Public Education Under Criticism" represents an attempt on the part of professional educators to consider the needs and the best ways of handling the criticisms of teachers, school administrators and teaching techniques. It makes an honest effort to let opponents and proponents present their particular attitudes for the purpose of appraising the credentials of our educational philosophy as well as practices. Selections made are presented under five sections, considered as areas of criticism and ways of handling them. The major criticisms include progressive education, fundamentals, religion, social studies, teacher education and teachers

PUBLICATIONS RECEIVED

Georgia Sachs Adams and Theodore L. Torgerson. *Measurement and Evaluation for the Secondary-School Teacher*. New York: The Dryden Press, 1956, pp. 658. $5.75.

Ake Bjerstedt. Interpretations of Sociometric Choice Status. Lund, Sweden: C. W. K. Gleerup, 1956, pp. 408 (paper). 40 sw. cr.

Edgar F. Borgatta and Henry J. Meyer, Editors. *Sociological Theory: Present-Day Sociology from the Past*. New York: Alfred A. Knopf, Inc., 1956, pp. 546. $6.75.

Daniel Brower and Lawrence E. Abt, Editors. *Progress in Clinical Psychology. Volume II*. New York: Grune & Stratton, 1956, pp. 364. $7.75.

Harrison Bullock. *Helping the Non-Reading Pupil in the Secondary School*. New York: Bureau of Publications, Teachers College, Columbia University, 1956, pp. 180. $3.75

D. Dee Castetter, Lloyd S. Standlee and Nicholas A. Fattu. *Teacher Effectiveness: An Annotated Bibliography*. Bloomington, Indiana: Institute of Educational Research, School of Education, Indiana University, 1954, pp. 105 (paper). $1.00.

The College Entrance Examination Board. Fifty-fourth Report of the Director. Princeton, New Jersey: College Entrance Board, 1955, pp. 107 (paper). $0.50.

Contributi del Laboratorio di Psicologia. Milano, Italia: Universita Cattolica del Sacro Cuore, Piazza S. Ambrogio, 9, 1955, pp. 544 (paper).

David G. Danskin. *Workbook in Educational Psychology*. New York: Henry Holt and Co., 1956, pp. 249. $2.75.

Ray Dotterer. *Postulates and Implications*. New York: Philosophical Library, Inc., 1955, pp. 509. $7.50.

Allen L. Edwards. *Statistical Methods for the Behavioral Science*. New York: Rinehart & Company, Inc., 1954, pp. 542. $6.50.

Vergilius Ferm, Editor. *Encyclopedia of Morals*. New York: Philosophical Library, Inc., 1956, pp. 682. $10.00.

Frieda Fromm-Reichmann and J. L. Moreno. *Progress in Psychotherapy*. New York: Grune & Stratton, Inc., 1956, pp. 352. $8.50.

J. Raymond Gerberich. *Specimen Objective Test Items: A Guide to Achievement Test Construction.* New York: Longmans, Green and Company, Inc., 1956, pp. 436. $4.75.

Leonard V. Gordon. *Gordon Personal Profile and Gordon Personal Inventory.* Yonkers-on Hudson: World Book Company, 1956 (paper). $0.50.

Andrew W. Halpin. *The Leadership Behavior of School Superintendents.* Columbus, Ohio: The Ohio State University Press, 1956, pp. 109 (paper). $2.00.

William E. Henry. *The Analysis of Fantasy: The Thematic Apperception Technique in the Study of Personality.* New York: John Wiley & Sons, Inc., 1956, pp. 305. $6.00.

Willliam A. Hunt. *The Clinical Psychologist.* Springfield, Illinois: Charles C. Thomas, 1956, pp. 207. $5.50.

A. M. Jordan. *Educational Psychology. Growth and Learning.* Fourth Edition. New York: Henry Holt and Co., 1956, pp. 600. $5.50.

Maurice H. Krout, Editor. *Psychology, Psychiatry, and the Public Interest.* Minneapolis: The University of Minnesota Press, 1956, pp. 217. $4.00.

Henry H. Linn, Editor. *School Business Administration.* New York: The Ronald Press Company, 1956, pp. 574. $7.50.

Richard Madden and Philip Peak. *Madden-Peak Arithmetic Computation Test.* Yonkers-on-Hudson, New York: World Book Company, 1956, pp. 18 (paper).

Stanley S. Marzolf. *Psychological Diagnosis and Counseling in the Schools.* New York: Henry Holt and Co., 1956, pp. 401. $4.00.

A. da Silva Mello. *Man: His Life, His Education, His Happiness.* New York: Philosophical Library, 1956, pp. 729. $6.00.

Henry Estill Moore. *Nine Help Themselves.* Austin, Texas: The University of Texas, 1955, pp. 286. (paper).

E. A. Peel. *The Psychological Basis of Education.* New York: Philosophical Library, Inc., 1956, pp. 303. $6.00.

Sandor Rado and George E. Daniels, Editors. *Changing Concepts of Psychoanalytic Medicine.* New York: Grune & Stratton, Inc., 1956, pp. 248. $6.75.

Anne Roe. *The Psychology of Occupations.* New York: John Wiley & Sons, Inc., 1956, pp. 340. $6.75.

DAGOBERT D. RUNES. *On the Nature of Man; An Essay in Primitive Philosophy*. New York: Philosophical Library, Inc., 1956, pp. 105. $3.00.

MORRIS S. SCHWARTZ and EMMY LANNING SHOCKLEY. *A Study in Interpersonal Relations*. New York: Russell Sage Foundation, 1956, pp. 289. $3.50.

F. ELEANOR SCHONELL. *Educating Spastic Children*. New York: Philosophical Library, Inc., 1956, pp. 242. $6.00.

GEORGE D. SPACHE and PAUL C. BERG. *The Art of Efficient Reading*. New York: The Macmillan Company, 1955, pp. 273 (paper). $3.00.

CATHERINE E. STELTZ. *As You See It: Twenty-four Pictures to Stimulate Group Discussion of Social and Moral Problems, with Suggestions for Their Use*. New York: Bureau of Publications, Teachers College, Columbia University, 1956, pp. 26 (paper). $2.95.

DAVID A. STEWART. *Preface to Empathy*. New York: Philosophical Library, Inc., 1956, pp. 157. $3.75.

PERCIVAL M. SYMONDS. *Dynamics of Psychotherapy: The Psychology of Personality Change. Volume I. Principles*. New York: Grune & Stratton, Inc., 1956, pp. 211. $5.50.

W. D. WALL. *Education and Mental Health: A Report Based upon the Work of a European Conference Called by UNESCO at the Musee Pedagogique in Paris*. New York: Columbia University Press, 1955, pp. 347. $3.00.

ROBERT W. WHITE. *The Abnormal Personality*. Second Edition. New York: The Ronald Press Company, 1956, pp. 644. $6.50.

J. WAYNE WRIGHTSTONE, JOSEPH JUSTMAN, MORRIS PINCUS and RUTH H. LOWE. *New York Test of Arithmetical Meanings. Level One and Two*. Yonkers-on-Hudson, New York: World Book Company, 1956, 48 pp. (paper).

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 48 April, 1957 Number 4

ANIMISTIC THINKING AMONG COLLEGE AND HIGH SCHOOL STUDENTS IN THE NEAR EAST

WAYNE DENNIS

Brooklyn College, Brooklyn, N. Y.

INTRODUCTION

In an earlier publication (*4*) the writer reported that approximately one-third of the students in several American colleges and universities attribute life to one or more inanimate objects. This finding has been corroborated by several other investigators (*1, 2, 6, 9*).

In comparison to many countries, the level of education in general, and the level of scientific knowledge in particular, are high in America. If animistic thinking is present in one-third of American college students, it seems likely that it is even more prevalent in countries in which the educational levels are lower, and in which natural science concepts and information are less widely disseminated than in the United States. The present paper reports data relative to this deduction, derived from college and high school students in the Near East.

PROCEDURE

Each subject was given a mimeographed sheet containing the following instructions:

"In connection with each of the objects listed below, indicate whether it is living or not living. Give a reason for each answer. Please answer the questions in the way that you think they would be answered by a biologist. For each object, the question is, is the object living in the same way that animals and plants are living?"

Below the instructions appeared the names of seven objects. For the subjects who were college students the objects were: a lighted match, the sun, the wind, the sea, lightning, a pearl and atoms. On

Bureau Ednl. Resear
Dated

the blanks given to high school students, "a river" was used instead of "the sea" and "petroleum" instead of "atoms." In the case of college students, the questions were presented and answered in English, which was their language of instruction. Since the teaching of the high school subjects was in Arabic, in their case the questions were presented and answered in that language.

SUBJECTS

The college students were enrolled in a woman's college and a university in Beirut, Lebanon. Although these institutions are located in Lebanon, their students come from many Near Eastern countries. In the university, only male students were used as subjects. The majority were students in the college of arts and sciences. High school data were obtained from three public high schools in Iraq. Two of these schools were located in Mosul and one in Baghdad. The latter school and one of the Mosul schools were day schools; the other was a night school attended primarily by adults who had discontinued their earlier education before the completion of high school.

In the colleges the animism questionnaire was given in the introductory psychology course. Since the majority of students at these institutions take this course, the sampling is believed to be representative of students at these institutions. In the high schools, the questionnaires were administered to first-, second- and third-year students in required courses, and consequently, the results are believed to be representative of these levels in these schools.

RESULTS

The number of subjects tested in each school and the per cent giving one or more animistic answers are indicated in Table I.

The most arresting fact about Table I is the generally high frequencies of animistic thinking found in all groups. The lowest figure (fifty-three per cent) was obtained for men in the university. All other percentages lie between seventy-four per cent and ninety-five per cent.

The difference between the college men and college women is probably due to differences in the institutions from which they come in regard to student selection and required science courses. Sex comparisons in the high schools, in which admission policies

TABLE I.—NUMBER AND PER CENT OF STUDENTS GIVING ANIMISTIC ANSWERS

	College Women	College Men	Baghdad Students	Mosul Day Students	Mosul Night Students	Total
Number of students	130	75	115	286	141	747
Per cent giving one or more animistic answers	77	53	95	81	74	79

and science courses for the two sexes are identical, indicate no sex differences. Comparisons of Christian and Moslem students, not included in Table I, indicate these two groups do not differ in frequency of animistic answers.

Table II indicates the frequency with which each object was said to be living by the members of each group. It will be noted that the lowest percentage obtained was for lightning. Among the university men only sixteen per cent stated that it was living. The university men yield a response frequency of twenty-one per cent for both the burning match and the wind. The highest per cents obtained were seventy-five per cent and seventy-eight per cent. In both cases these refer to the pearl, and the scores were made by the

TABLE II.—PER CENT OF RESPONSES TO EACH QUESTION WHICH WERE ANIMISTIC

	College Women	College Men	Baghdad Students	Mosul Day Students	Mosul Night Students
Match	47	21	50	27	42
Sun	38	26	66	36	35
Wind	34	21	56	24	35
Sea	41	24			
Lightning	30	16	32	23	25
Pearl	26	24	78	75	42
Atoms	55	31			
River			54	28	42
Petroleum			49	23	40
Per cent of total answers animistic	39	23	55	34	37

TABLE III.—REASONS GIVEN FOR ANIMISTIC ANSWERS

	College Women	Baghdad Students
	Per cent	*Per cent*
Moves, changes, does something	39	38
Supports or produces life	21	7
Has a beginning and an end	14	4
Grows	16	23
Uses oxygen, produces energy	6	16
Miscellaneous	4	12

Iraqi high school groups. It will be noted that in four of the five groups one-third or more of the total responses were animistic.

In the case of two groups, the college women and the Baghdad students, the reasons given for positive answers have been examined and classified. The results are shown in Table III. It will be noted that in the two groups, respectively, thirty-nine per cent and thirty-eight per cent of the answers are what Piaget would call Stage II or Stage III answers: the object is living because it moves, either by itself or from some other cause. These Piaget classifies as non-adult answers. While these answers occur frequently in our group, the simplest sort of answer obtained from children, namely that an object is living because it is useful, was rarely encountered among college and high school students. Informal studies of Lebanese elementary school pupils which will not be reported in detail show that Stage I answers (answers in term of use) predominate at the earlier age levels in Lebanon, as they do elsewhere (*3, 8*).

The statement that an object is living because it grows should perhaps not be listed as an instance of wrong reasoning. The reasoning is not incorrect; only the facts are wrong. Among the Iraqi subjects this answer was often given in respect to the pearl.

The statements that an object, such as the burning match, is living because it is using oxygen illustrate the fact that science education does not always operate to reduce animistic answers. Apparently the students had learned that living things use oxygen, and they reasoned that whatever uses oxygen, must be living.

College and high school students give two types of answers which are seldom given by children. The answers occur among American college students (4) as well as in the present sample. One is that the sun, or the sea, or some other object is living because it produces, or is necessary for, life. This answer seems to result from the premise that like produces like, and therefore, that whatever is responsible for, or necessary for life, must itself be living. The second notion is that since life has a beginning and an end, all things which have a beginning and an end, as do a flame and a flash of lightning, are therefore living. These answers have not been reported by Piaget in his four stages of animism, presumably because he questioned only children. They represent categories or "stages" in excess of his four.

DISCUSSION

It is obvious from the present study, as well as from earlier ones (*3*, *4*), that Piaget's assumption that animism is limited to children is incorrect. It seems probable that a degree of animistic thinking can be found in some adults in any society. The actual percentage of animistic replies will differ from question to question and from group to group.

While the cultural area in which the present study was conducted was a center of one of the earliest civilizations, extensive modern educational programs have been introduced only recently. Science education, as compared to linguistic, literary and historical instruction, is undoubtedly still weak. These two facts probably account in the main for the high incidence of animistic replies to our questions.

We do not believe that the prevalence of animistic thinking in our subjects is due to positive cultural influences. Christianity and Islam, to which our subjects belong, both reject the view that objects of nature and forces of nature are themselves alive. Both of these religions would reject such ideas as tending toward idolatry.

We have indicated earlier (*3*) our belief that animism arises autogenously in individuals who lack an adequate background in modern biological concepts and knowledge. The persistence of animism into the adult years, in the Near East, is to be accounted for in terms of educational handicaps rather than in terms of a positive animistic tradition.

SUMMARY

Seven hundred and forty-seven Near Eastern college and high school students were questioned in regard to whether each of sev-

eral objects is living or not living. Seventy-nine per cent gave one or more animistic answers. In the various groups the frequencies of animistic answers were much higher than those obtained among American students of comparable educational placement. The differences are presumed to reflect the differences in diffusion of scientific concepts and information in these populations.

ACKNOWLEDGMENT

The writer is indebted to Miss Suad Wakim, Mrs. Ada Khaza'al and Mr. Abdul-Kadir Hubayti for obtaining the student responses on which the present paper is based, and to Mrs. Yvonne Sayegh for analysis of the data. The study was conducted in 1955–56 when the author was a visiting professor at the American University of Beirut, Lebanon. Expenses of the study were defrayed by a grant to the American University by the Rockefeller Brothers Fund.

REFERENCES

1) C. R. Bell, "Additional Data on Animistic Thinking," *Scientific Monthly*, 79: 67–69, 1954.

2) C. N. Crannell, "Responses of College Students to a Questionnaire on Animistic thinking," *Scientific Monthly*, 79: 54–56, 1954.

3) W. Dennis, "Animism and Related Tendencies in Hopi Children," *Journal of Abnormal and Social Psychology*, 38: 21–36, 1943.

4) W. Dennis, "Animistic Thinking Among College and University Students," *Scientific Monthly*, 76: 247–249, 1953.

5) W. Dennis and R. W. Russell, "Piaget's Questions Applied to Zuni Children." *Child Development*, 11: 181–187, 1940.

6) D. C. Lowrie, "Additional Data on Animistic Thinking," *Scientific Monthly*, 79: 69–70, 1954.

7) J. Piaget, *The Child's Conception of the World*, New York, Harcourt, Brace, 1930.

8) R. W. Russell, "Studies in Animism; II. The Development of Animism," *Journal of Genetic Psychology*, 56: 353–366, 1940.

9) V. Voeks, "Sources of Apparent Animism in Students," *Scientific Monthly*, 79: 406–407, 1954.

EFFECTS OF PRAISE AND REPROOF ON READING GROWTH IN A NON-LABORATORY CLASSROOM SETTING

HARRY F. SILBERMAN

Division of Teacher Education

Municipal Colleges of the City of New York

This report compares teachers whose verbal behavior is characterized by different amounts and combinations of praise and reproof. Praise and reproof are verbal incentives that have been found to enhance performance on learning tasks (*1, 3, 8, 9*). In some experiments praise has been superior to reproof in its effect upon performance (*1*). In other experiments, reproof has been shown to be more effective (*2*). Three sources of variation which might account for the conflicting results obtained in these studies are: (1) school or teacher differences, (2) motivational differences among pupils engendered by novel laboratory settings, and (3) differences in operational definitions of praise and reproof.

The present study, using forty-nine teachers in grades three to six in nineteen schools, is designed to study the effects of praise and reproof in teachers' verbal behavior on reading growth of pupils in non-laboratory classroom settings. It was the intention in this study to investigate not only praise and reproof but the combination of the two. The combination of praise and reproof is obtained by multiplying the praise scores by the reproof scores. A convenient way of interpreting this product variable is provided by the following table:

		PRAISE	
		hi	lo
REPROOF	hi	a	b
	lo	c	d

This table shows four possible combinations of praise and reproof in teachers' verbal behavior. If the PR variable, (praise-reproof combination), is significantly related to the reading growth criterion, teachers in the "a" quadrant differ from teachers in the "d" quadrant in promoting reading growth provided other relevant factors have been controlled.

Two additional variables, teachers' verbal output as measured by the number of statements teachers make and the relative amount of time spent on reading skills, were also chosen for study.

The hypothesis of this study, stated in the null form, is that measured reading growth is unrelated to a weighted combination of the relative amounts of praise, reproof, combined praise by reproof, verbal output, and the observed amount of time spent on reading skills. If this hypothesis is rejected, the alternatives are that either praise, reproof, praise by reproof, verbal output, time on reading skills, or some combination of these five independent variables, is related to measured reading growth of pupils.

PROCEDURE

Sample. This study was conducted during the school year 1954–55 using twenty-three third-grade classes, thirteen fourth-grade classes, nine fifth-grade classes and four sixth-grade classes located within nineteen New York City public elementary schools. Test data were collected from a total of nine hundred and four pupils. All forty-nine teachers were in their first year of teaching. They constitute a sub-sample drawn from sixteen hundred and twenty-eight members of the student teaching classes of 1953–54 at the municipal colleges of New York City: City, Hunter, Brooklyn and Queens. Descriptive information about the sixteen hundred and twenty-eight student teachers from whom the forty-nine teachers in this study were drawn is given in a previous publication (*6*).

Classroom observations. Six classroom observers were employed, each observer paying each of the forty-nine teachers two half-hour visits during the school year. Each teacher was visited by only one observer at a time. An observational schedule and rating form, details of which are given in a previous report (*5*) was developed for use by the classroom observers. The dimensions in the observation schedule with which this study is concerned are called "expressive behaviors" and "subject-matter emphasis."

The "expressive behaviors" dimension, adapted from Withall

(*11*), consists of a group of seven spaces on an observation card for tallying teachers' verbal behavior and certain non-verbal behaviors. The half-hour visits were divided into six five-minute observation periods. During the second, fourth, and sixth five-minute observation period the observers tallied every remark the teacher made in one of the following categories of the "expressive behaviors" dimension.

(1) *Supportive statements.* Teachers' remarks which trained observers construe as approval of the pupils or of their responses are tallied in this portion of the card. For example, comments such as "Right," "Good," "Fine work," which inform the pupil that he or his responses adequately satisfy the standards which the teacher expects the pupil to meet, are classified as *praise.*

(2) *Problem-structuring statements.* Statements in which the pupil is offered facts, ideas, opinions, or questions in an objective non-threatening manner.

(3) *Neutral statements.* Statements not classified in any other category.

(4) *Directive statements.* Statements that limit the pupils' choice of behaviors. Examples are: "Answer the questions on page 12" and "Pass the papers forward."

(5) *Reproving statements.* Tallies of hostile, disapproving verbal behavior which the observers construe as threatening to the pupil. "You're all wrong" and "Won't you ever learn?" are illustrations of reproving statements.

During the first series of classroom visits the six observers made a new tally on this dimension only when the teacher's remarks shifted from one category to another. During the second series of classroom visits the six observers tallied the frequency of occurrence of statements in each category.

The "subject-matter" dimension of the observation schedule is a group of ten categories for tallying the subject areas in which the activity of the class is centered. One tally was made in this section by each observer during every five-minute portion of the half-hour observation period. The subject areas are: reading, mathematics, language arts, social studies, science, recreation, arts and crafts, music, social processes and test periods.

The first series of observations was arranged during February and March of 1955. Teachers were assured on several occasions that the observer was simply recording what went on in the classrooms and was in no way judging or rating the teachers. They were told of the purely scientific interest of the investigators and were repeatedly assured that all information obtained was entirely confidential. The teachers were notified in advance of the date and approximate time of each visit, and sometime near the appointed time an observer unobtrusively entered the classroom, taking a seat at the rear of the room. The observer proceeded to observe and place tallies on the observation schedule card reserved for that teacher, employing a stop-watch to time the five-minute portions of the half-hour visit. The second series of visits was arranged during March and April, 1955.

The criterion (dependent variable). A detailed discussion of the development of the reading growth criterion is presented in another report (7) in the current series of research studies on teacher behavior being performed by the Division of Teacher Education of the College of the City of New York (*10*). Reading ability of the pupils in all forty-nine classes was measured during November 1954 and again during April or May of 1955 using the California Reading Test (Elementary): Word Form, Word Recognition, Meaning of Similarities, and Interpretation of Meanings. Form S of the California Test of Mental Maturity, Non-Language Section, was also administered during November, 1954. The reading growth indices to be used as a criterion in this study consist of the means of the final reading scores of the pupils in the teachers' classes, adjusted by covariance control for the mental maturity scores and the initial reading scores. This procedure effectively equates pupils with respect to mental maturity and previous achievement in reading. In the analysis of the data these reading indices were converted to normalized "T" scores.

Composition of the independent variables. This study entailed the use of six variables, consisting of one dependent variable described above and five independent variables.

The five independent variables are operationally defined as follows:

(1) *Praise.* Total number of supportive statements tallied.

(2) *Reproof.* Adjusted total number of reproof statements tallied.

(3) *Praise by Reproof.* A composite of praise and reproof consisting of the product of the number of supportive and adjusted number of reproof statements tallied. The highest "PR" score would be obtained by a teacher whose verbal behavior is characterized by the largest amounts of praise and reproof in combination.

(4) *Verbal Output.* Total number of verbal statements tallied in the "expressive behaviors" dimension of the observation schedule for each teacher.

This is an estimate of the relative amount of classroom verbal behavior exhibited by each teacher.

(5) *Reading Time.* Total number of tallies for each teacher in the reading section of the "subject-matter" dimension of the observation card. This constitutes an estimate of the relative emphasis or amount of time that each teacher devoted to specific instruction in reading.

Preliminary analysis of the five independent variables indicated that the 49 teachers did not rank differently on the two series of observations on all but the reproof scales. As a result, the number of tallies on the two series were added for each teacher on all but the reproof scales. The reproof measures obtained on each of the two series of observations were given equal weight by converting the measures on each series to standard scores and then multiplying the derived scores by the standard deviation of the series with the largest variance. The obtained scores were then combined to form the reproof score for each teacher.

In the analysis of the data all variates were converted to normalized "T" scores. Differences among teachers were significant at the 0.01 level on each of the six variables. The reliabilities of these variables when estimated by analysis of variance (*2, 7*), are shown in Table I.

RESULTS AND DISCUSSION

The intercorrelations among the independent variables in this study are shown in Table II. Both praise and reproof are significantly related to total verbal output.

The relationships between the reading growth scores and the independent variables is shown in Table III. The multiple correlation is not significant at the 0.05 level; hence, this sample does not contradict the hypothesis that in the population there is no correlation between reading growth and a weighted combination of the five independent variables. In other words, the independent variables do not predict the reading growth of pupils in this sample.

TABLE I.—ESTIMATED RELIABILITIES OF EXPERIMENTAL VARIABLES

Reading growth	0.87
Praise	0.71
Reproof	0.84
Verbal Output	0.80
Reading Time	0.64

Bur Ednl.

TABLE II.—INTERCORRELATIONS OF FIVE INDEPENDENT VARIABLES

	Praise	Reproof	Praise by Reproof	Verbal Output	Reading Time
Praise		−.1605	+.6370**	+.6057**	+.1839
Reproof			+.6274**	+.4304**	−.0036
Praise by reproof				+.7973**	+.2050
Verbal output					+.2745
Reading time					

** Significant at the 0.01 level.

TABLE III.—RELATIONSHIPS BETWEEN READING GROWTH SCORES AND FIVE INDEPENDENT VARIABLES

Independent Variables	Zero-Order Correlation	Partial Regression Coefficients (Beta Weights)	Standard Errors
Praise	+.0258	+.2806	.5657
Reproof	−.1539	+.1125	.5597
Praise by reproof	−.1282	−.4522	.7075
Verbal Output	−.0519	+.1432	.2574
Reading Time	−.1951	−.1929	.1616

Multiple Correlation $R_{y.12345} = .2795$.

In view of the non-significant relationships between the criterion and the independent variables employed in this study, no definite conclusions can be drawn. Whether or not an increase in the number of teachers observed, or other precautions designed to enhance the precision of the experiment, would have produced different findings is a matter of conjecture. Perhaps the observed teacher behaviors were atypical due to the presence of the observers in the classroom. If such were the case, the biased observations would have produced independent variables that are not representative of the teacher, and the negative results would hardly be surprising. A methodological limitation which should be noted in this study was the inability to assign pupils to the forty-nine classrooms at random. Differences among teachers on the independent variables might consequently be attributed to differences in the pupils assigned to those teachers. A teacher who issues large amounts of reproof, for instance, might argue that she is the victim of an emotionally disturbed class and, given a better class, would certainly produce a more permissive climate and more growth in reading. These are problems for further research.

The results of this study indicate that if there is a relationship between the verbal incentives and learning to read, it is a low relationship. If we accept this interpretation, certain implications for further exploration are evident. Many educators extoll praise and mild reproof as classroom incentives and offer a carte blanche endorsement of both as if praise and reproof contained some inherent merit. Yet, praise and reproof, as defined in this investigation, did not have appreciable effects on reading growth. Perhaps classroom situational variables not observed in this study are much more important to pupil growth than praise and reproof per se. The nature of the situations in which the incentives were employed and the nature of the pupils receiving the incentives may have been much more important variables to observe. At any rate, in this study praising and reproving remarks were tallied whether or not the incentives were contingent upon the correctness of pupils reading responses. The teachers verbal behavior during the teaching of the criterion skill as well as during the teaching of other skills was used.

Granting these limitations, the present study does not justify a blank endorsement for the use of praise or reproof per se in the classroom, and points up the need for the verification of laboratory experiments by observational studies of normal classroom behavior.

SUMMARY

Forty-nine beginning teachers were each visited twelve times during the 1954–55 school year and their verbal behavior was categorized and tallied. Pupils were given reading tests at the beginning and end of the school year.

A multiple correlation was computed between reading growth and five independent variables, namely, praise, reproof, praise by reproof, verbal output, and time devoted to reading skills. Results of the analysis did not yield significant relationships between the independent variables and reading growth of pupils.

A discussion of the limitations of the study and of possibilities for further research on this problem was presented.

REFERENCES

1) Benjamin Brenner, *Effect of Immediate and Delayed Praise and Blame upon Learning and Recall.* Contributions to Education, No. 620, Teachers College, 1934, 52 pp.

2) C. J. Hoyt, "Test Reliability Estimated by Analysis of Variance." *Psychometrika*, 1941, 6, pp. 153–160.

3) Elizabeth B. Hurlock, "An Evaluation of Certain Incentives used in School Work." *Journal of Ed. Psych.*, 16: 145–159; 1925.

4) Tatania Juzak, "The Effects of Praise and Reproof on the Generalization of Learned Concepts." *Journal of Psychology,* 1955, 39: 329–340.

5) D. M. Medley and H. E. Mitzel, *Studies of Teacher Behavior: The Refinement of Two Techniques for Observing Teachers' Classroom Behaviors.* Research Series, No. 18, Division of Teacher Education, Board of Higher Education of the City of New York, October, 1955.

6) H. E. Mitzel, et al. *Studies of Teacher Behavior: Descriptive Notes on a Sample of Student Teachers.* Research Series No. 27, Division of Teacher Education, Board of Higher Education of the City of New York, July, 1955.

7) H. E. Mitzel, and D. M. Medley, *Studies of Teacher Behavior: Pupil Growth in Reading—An Index of Effective Teaching.* Research Series No. 32, Division of Teacher Education, Board of Higher Education of the City of New York, May, 1956.

8) H. O. Schmidt, *The Effects of Praise and Blame as Incentives to Learning.* Psychological Monographs, 53, No. 3, 1942, 56 pp.

9) George G. Thompson and C. W. Hunnicut, "Effects of Repeated Praise or Blame on the Work Achievement of Introverts and Extroverts." *Journal of Ed. Psych.*, 35: 1944, pp. 257–266.

10) E. Wandt and H. E. Mitzel, *Studies of Teacher Behavior: Plan for a Program of Research.* Research Series No. 21, Division of Teacher Education, Board of Higher Education of the City of New York, June, 1954.

11) J. G. Withall, "The Development of a Technique for the Measurement of Social-emotional Climate in the Classroom." *Journal of Exp. Educ.*, 1949, 17, pp. 347–361.

EFFECT OF STEREOTYPED ATTITUDES ON LEARNING

ROBERT E. EGNER and ALVAN J. OBELSKY

One of the most recent and promising of the new developments in educational testing has been the work done in measuring stereotyped conditioning in students. Students arriving in college are hardly the "tabula rasa," attitudinally speaking, that college educators seem to have long taken for granted. Each has been subject to a unique schooling of attitude conditioning in the family, church, neighborhood, and in pre-college education. Some have relatively fixed and naïve opinions on topics about which they are usually not yet competent to express an opinion; others are more flexible and sophisticated and "liberal" in their attitudes. It is this sort of psychological trait that stereotypy testing is designed to measure.

The development of the stereotypy syndrome and its measurement are of quite recent origin. The monumental study, *The Authoritarian Personality,*[1] was the first large-scale and systematic attempt to evaluate factors concerned with stereotyped conditioning, particularly where it concerned racial discrimination. This work, both because of the importance of its subject content and its methodological analysis, stimulated further work in other phases of stereotyped conditioning, most noticeably as it applied to political attitudes.[2] Its implication for the educational field, however, had not been missed and within the past five years a considerable amount of work, primarily in the development of valid tests of stereotypy, has been done. It is interesting to note that the first public presentation dealing with the relevance of stereotypy to the learning process dates only from 1952,[3] while the earliest devised test for measuring stereotypy in students, that of the American Council on Education, appeared in 1951. Since that time "Inventory of Beliefs" tests, essentially adaptations of the American Council on Education test, have been used at a number of institutions, notably at the University of Chicago. During the past three years at Northland College the Form T test (used at the University of Chicago) has been given to students in a required junior-year philosophy course by Professor R. E. Egner, and the combined results analyzed and significant correlations examined. The results of these efforts will be briefly described in the following paragraphs.

---

[1] T. W. Adorno, et al., *The Authoritarian Personality* (Harper, 1950). Credit, of course should be given also to the earlier studies relating to the theory and measure of personality syndromes conducted by Henry Murray and his associates at the Harvard Psychological Clinic. H. Murray, et al., *Explorations in Personality,* (Oxford, 1938).

[2] See especially: *Studies in the Scope and Method of the "Authoritarian Personality,"* edited by Chrisite and Jahoda (Harper, 1950). This book provides a critique of *The Authoritarian Personality* and, in addition, develops some interesting modifications and extensions of ideas contained in the earlier work.

The Form T test used in this study is composed of an inventory of one hundred beliefs, of which sixty are common stereotyped expressions. The examinee is asked to indicate quickly his acceptance or rejection of the statements. No way is provided for indicating a neutral position and all items must be answered. The test is designed to indicate only the *direction* in which the personality has been conditioned to move. The stereotyped personality will show a marked tendency to accept the stereotyped beliefs and to reject the non-stereotyped statements; with the non-stereotyped personality, the two tendencies are reversed. A sample of the beliefs taken from the Form T test are shown below together with the appropriate key.

(A) I strongly agree or accept the statement.
(B) I tend to agree or accept the statement.
(C) I strongly disagree or reject the statement.
(D) I tend to disagree or reject the statement.

(1) Literature should not question the basic moral concepts of society.

(2) Any man can find a job if he really wants to work.

(3) Science will eventually explain the origin of life.

(4) No task is too great or too difficult when we know that God is on our side.

---

[3] S. Goldberg, and G. Stern, "The Authoritarian Personality and General Education." Paper presented at the 60th Annual Convention of the American Psychological Association, September, 1952, Washington, D. C. Abstracted. A recent and more complete study dealing with problems of stereotyped behavior in college students may be found in chapter 10 of *Methods of Personality Assessment* by Stern, Stein and Bloom (Free Press, Glencoe, 1956).

(5) A sexual pervert is an insult to humanity and should be punished severely.

(6) The twentieth century has not had leaders with the vision and capacity of the founders of this country.

(7) Other countries don't appreciate as much as they should all the help that America has given them.

(8) No censorship on the presumed morality of books and movies can be justified.

(9) Our rising divorce rate is a sign that we should return to the values which our grandparents held.

(10) Most intellectuals would be lost if they had to make a living in the realistic world of business.

This test was administered over a period of three years to a total of two hundred and eighteen students enrolled in an introductory philosophy course (junior level) which was required of all students at the college. The samples upon which the subsequent analysis is based represent students from this total who scored at the extremes on the test and who were also equated for intelligence.[4] Twenty low and twenty high scorers, composing the two samples, were selected for further investigation. The problem was to determine if any significant correlation existed between high and low scoring on the stereotypy test and performance in the various subject areas. If one interprets the significance of stereotyped conditioning on learning ability in terms of the number of successful students in each sample the differences in performance between the two groups readily becomes apparent. The figures in parentheses indicate the number of students in each sample receiving a grade of "C" or above in the various areas of learning.

Humanities	Stereotype (12)	Non-Stereotype (19)
Social Science	Stereotype (11)	Non-Stereotype (17)
Biological Science	Stereotype (18)	Non-Stereotype (17)
Mathematics	Stereotype (17)	Non-Stereotype (12)
Natural Science	Stereotype (16)	Non-Stereotype (12)

It is clear that the non-stereotypes performed appreciably better in the Humanities and in the Social Science subjects, while in Mathematics and in the Natural Sciences the performance of the stereotypes was superior. In the Biological Sciences both groups did about equally well.[5] In view of the observation that a higher percentage of stereotypes tends to drop out in the first two years of college it is quite possible that a selective element is present in the sampling. Any adjustment for this, of course, would improve the showing of the non-stereotype group.

---

[4] In both samples students were selected whose performance on the A.C.E. Psychological test were approximately the same—individual scores were located within a range of eight centiles.

The above results, in spite of the small sampling taken, are closely in line with the observations made from similar studies at other institutions. The student most likely to succeed in courses which involve relatively high levels of flexibility of outlook is the non-stereotype. As one moves from the concrete to the abstract a larger percentage of stereotyped students begins to manifest evidence of frustration, culminating in more frequent academic failure, once their threshold of tolerance has been exceeded. Figures would most likely underestimate their rate of failure since a substantially larger number of stereotypes withdraw from college before eventual failure. According to the University of Chicago examiner, Dr. Benjamin S. Bloom, about one fourth of the stereotyped students drop out of the College of General Education by the end of their first quarter.[6] The reasons given for these withdrawals provide a clue to the kind of atmosphere in which the stereotype thrives. Some of the most common follow: The education was not very practical; assignments were never very definite; the discussion technique represented a poor educational method (apparently because it lacked an authoritarian teacher); definite and final answers to problems in the humanities and the social sciences were never given.

An educator who is reasonably alert to the performances of his individual students will have little difficulty in recognizing the presence of stereotyped attitudes and opinions among his students. Evidence can be found in abundance in written examinations, particularly in the essay type where the student enjoys more or less free reign in putting down his thoughts. Oral responses to questions and private discussions with the instructor—

---

[5] In Humanities and Social Science the differences beween the groups are significant at the 5% level; in Mathematics the difference is just below statistical significance.

[6] From an address delivered at the Workshop in Higher Education, North Central Association Study of Liberal Arts Education. University of Chicago. Summer, 1953.

indeed, any form of communication—will reflect these attitudes. As long as the stereotype is communicative, and it is probably the case that as a group they are more inclined to be so, his characteristics can be readily observed. The "true" stereotype is entirely convinced, at least on an emotional plane, of the correctness of his views and so does not hesitate to express them. This is true of the "left" and "right" stereotypes; both insist on the correctness of their respective positions which are located at the extremes of the continuum line. To take an example from the economics class, a "left" stereotyped student will tend to view the capitalist system as one characterized by a highly unequal distribution of income and unequal economic opportunities for its members, and exploitation of the working classes by those in economic power. Those on the "right" consider the institution of private property a sacred mainstay of freedom, both economic and political, with inequalities a necessary driving force in the economy, and view the influence of labor as baleful to freedom and restrictive of the natural working of the economic system. It would not be difficult to find comparable examples of "left" and "right" stereotypy in the other academic disciplines.

This "black" and "white" mentality of the stereotyped student can be so easily recognized (at least by an instructor who does not share the same species of stereotyped attitude) that a question very quickly arises concerning the usefulness of an elaborate test to detect its presence. The test clearly is of greatest use where large numbers of students are involved whose attitudes are as yet unknown. Under these circumstances it is convenient and, at the same time, it permits a rough quantitative appraisal of the degree of stereotypy present. At a large institution it could be given as part of a battery of tests and, if desired, correlations among performances on the stereotypy and other tests could be readily found. In the smaller school, where classes are generally limited and contact between students and instructor more intimate, the value of the test is considerably diminished.

An area which has received little attention is the matter of "de-stereotyping" the stereotyped student. One might think that this is a task for the psychological advisor, yet changing fixed and inadequate beliefs of students is a constant function of education. The extreme stereotype who holds to irrational beliefs with an obsessive force might, perhaps, be put down as a case in pathology. But the more common stereotype presents a problem for the educator which differs only in degree from that which is faced in the ordinary affairs of education. Viewed in this light, the stereotype becomes simply a more intractable case to handle, though still amenable to the usual techniques of effective teaching. "Bare" facts, served in imposing dosages, cannot be ignored by the student for very long, and if the patience of the instructor is not lacking it should be a matter of time before a strongly entrenched stereotyped attitude begins to break down. Removal of the student from the emotional environment fostering the stereotyped conditioning would, of course, contribute a great deal to the progress.

The entire problem of stereotypy has such broad implications for society as a whole that it seems superfluous to attempt to offer further justification for our concern with this subject. Studies of the authoritarian personality, which deal with a rather specific form of stereotypy, give adequate evidence that the fate of our democratic institutions themselves is at stake in this issue. The work now being done on stereotypy and the learning process should convince educators that they are in a more potent position to advance the cause of democratic society than might have appeared.

LOGICAL VERSUS EMPIRICAL SCORING KEYS: THE CASE OF THE MTAI

N. L. GAGE

Bureau of Educational Research

College of Education, University of Illinois

The Minnesota Teacher Attitude Inventory (MTAI) (*1*), has scoring weights which are non-monotonically related to degree of agreement with the item. This is true of ninety of the MTAI's one hundred fifty items. That is, the same scoring weight is assigned to responses on both sides of the "undecided" position, or different scoring weights are assigned to responses on the same side of the "undecided" position. Such items can be identified wherever the 0 weight falls somewhere other than between the +1 or −1 weights, or wherever two or more +1 or −1 weights are in non-adjacent alternatives. (Non-monotonic weights would not be so puzzling if they proceeded +1, −1, −1, −1, +1, or −1, +1, +1, +1, −1. There are, however, no items like this in the MTAI.)

For example, consider Item 1, "Most children are obedient." The scoring weights for this item are SA, +1; A, −1; U, O; D, −1; SD, 0. It is difficult to make psychological sense out of such scoring weights. The general direction or trend of the weights for all items does indeed assign positive weights to a permissive, child-centered orientation. But the non-monotonic weights defy any such conceptualization. And they imply that dimensions other than degree of agreement with the items are relevant to the test's purpose.

Two defenses might be offered for these non-monotonic weights: (a) they tend to defeat attempts at faking; (b) the scoring weights are based on empirically determined differences between the responses of one hundred very good and one hundred very poor teachers.

Both these defenses can be pierced by available evidence. Thus, the MTAI has indeed been found to be fakable* (*4*), despite the complex and apparently illogical scoring weights of many of the items. Second, experience has shown (*3*, p. 450) that very large samples (of the order of five hundred cases) are needed before empirical scoring weights become stable. Hence, it would be surprising if the MTAI's non-monotonic weights, based on only one hundred teachers in each extreme group, were not merely reflecting random fluctuations due to sampling error.

---

* At least in one sense.

We can therefore hypothesize that the non-monotonic, empirically derived scoring weights have lower validity than a logical set of scoring weights. To test this hypothesis, two "logical" scoring keys were made: (a) a dichotomous-logical key in which both choices on a given side of "undecided" were weighted +1, and the remaining three choices were weighted 0; (b) a trichotomous-logical key, in which the same choices as in *a* were weighted +1, but the two choices on the other side of "undecided" were weighted −1. In each instance we weighted +1 that side of the response continuum which contained the +1 weight in the published key. The following item illustrates the three keys whose validities could be compared:

1. Most children are obedient.	SA	A	U	D	SD
Published key	+1	−1	0	−1	0
Dichotomous-logical key	+1	+1	0	0	0
Trichotomous-logical key	+1	+1	0	−1	−1

We used data obtained in another study (*2*) involving ninety-seven of the ninety-eight teachers and all their twenty-seven hundred pupils in grades four to six of a midwestern city. Our criterion was the same as that used, in part, by the authors of the MTAI. This was the mean of the ratings given the teacher by her pupils on Leeds' "My Teacher" rating scale.

Table I shows that the split-half reliabilities of the scores obtained with the two "logical" keys were higher (0.93 and 0.94) than that of scores obtained with the empirical key (0.90). Hence, we do not need to depend on the empirical key to measure something with the MTAI that is consistent over items.

The validity coefficient of the trichotomous-logical key was higher than that of the published empirical key: 0.28 as against 0.26. The important point is not that the logical key had a slightly higher validity. Rather it is noteworthy that it did *not* have lower validity.

These findings have both practical and theoretical importance. In future use of the MTAI, the published empirical key should be

Table I.—Reliability and Validity of Three Types of MTAI Score

N = 97

	M	S.D.	r_{tt}	r with Criterion
Criterion (Mean rating by pupils on "My Teacher")	20.7	10.4	.91*	
Published empirical key for MTAI	56.7	35.5	.90**	.26
Dichotomous-logical key for MTAI	111.8	17.1	.93**	.23
Trichotomous-logical key for MTAI	88.2	28.9	.94**	.28

* Horst (*4*).

** Corrected split-half.

reëvaulated in comparison with our "logical" key; it seems probable that the empirical validity of the MTAI can be improved by using a logical scoring key.

Of even greater potential significance is the implication that what the MTAI measures can be conceptualized in terms of relatively simple and communicable theory. If we can abandon the complex and arbitrary-seeming published scoring weights for the items, the scoring key becomes readily comprehensible and predictable. Thus, we have found that three graduate students in educational psychology—instructed to assign scoring weights in accordance with a permissive, child-centered, non-authoritarian, and emotionally secure orientation—were able to predict the logical scoring weights unerringly.

SUMMARY

The Minnesota Teacher Attitude Inventory had slightly higher reliability and validity when scored with a logical key than when scored with the published empirically derived scoring weights, which are non-monotonic for ninety of the one hundred fifty items. Practically, this suggests reëxamination of the published key. Theoretically, this means that the MTAI's meaning becomes relatively easy to formulate.

REFERENCES

1) W. W. Cook, C. H. Leeds, and R. Callis, *Minnesota Teacher Attitude Inventory,* New York, Psychological Corp., 1951.

2) G. M. Della Piana, and N. L. Gage, "Pupils' Values and the Validity of the Minnesota Teacher Attitude Inventory," *J. Educ. Psychol.*, 46: 167–178, 1955.

3) J. P. Guilford, *Psychometric methods*, second edition, New York, McGraw-Hill. 1954.

4) P. Horst, "A Generalized Expression for the Reliability of Measures," *Psychometrika*, 14: 21–31, 1949.

5) W. Rabinowitz, "The Fakability of the Minnesota Teacher Attitude Inventory," *Educ. Psychol. Measmt.*, 14: 657–664, 1954.

STABILITY OF INTRA-INDIVIDUAL PATTERNING OF MEASURES OF ADJUSTMENT DURING ADOLESCENCE

FRED T. TYLER

University of California

INTRODUCTION

According to certain theoretical positions, as indicated in various textbooks, there is a high degree of constancy and consistency in much of the individual's behavior. Empirical evidence on this point is relatively scarce since problems related to the behavior of the individual over a period of time have received much less attention than have those dealing with the behavior of groups of individuals at some specific point of time. Yet understanding and prediction of an individual's behavior depend upon knowledge concerning stability and consistency of his behavior: stability in the behavior of a group of subjects is no guarantee of stability in each member of the group.

The importance of questions dealing with the stability of personality is apparent: prediction is feasible if behavior is consistent over a period of time. According to Olson, prediction is possible when growth is continuous: "The fact that there is continuity to the growth process makes it possible to project curves and to predict" (*6,* p. 36). However, the present writer believes that prediction depends more upon stability and consistency than upon continuity of growth, since, in our opinion, growth may be continuous even while being irregular and erratic. Olson appears to accept the concept of stability of personality: "The stability of personality results in a strong resistance to displacement by temporary environmental variations" (*6,* p. 276).

ANALYSIS OF THE INDIVIDUAL

Analyses of physical and behavioral data of the single case have been, in the past, primarily graphical in nature, although other procedures are beginning to appear. Kerlinger (*5*) employed an analysis of variance technique to investigate the validity of the principle of unity of growth, or the concept of "homogeneous" and "heterogeneous growers," to use his terms. However, the "subtractive" scores he used in his analysis do not seem appropriate since consideration of his methodology shows that the use of "subtractive" scores may indicate the presence of heterogeneity of growth when in fact the *growth ages* imply homogeneity; and vice versa. Tyler (*9*) illustrated correlational [P-technique (1)] and factorial analyses of physical data for the individual subject. The effects of "staggering" longitudinal data to take into account the fact that growth in different physical characteristics is sequential rather than synchronous in nature are also under investigation by means of correlational analyses. Analysis of variance of a variety of behavioral data expressed in standard scores has also been used to investigate the validity of the principle of unity of growth.[1]

Spearman's *rho* and Kendall's *W* also appear to offer possibilities for the statistical analysis of data from the individual case. The present writer recently reported, for a series of behavioral measures, the values of rho between the measures obtained at two ages in the adolescent period for each of sixteen boys.[2] These rhos were 0.74, 0.71, 0.67, 0.63, 0.60, 0.54, 0.52, 0.46, 0.44, 0.42, 0.33, 0.09, 0.08, and −0.42, indicating a tendency for some similarity in the patterning of the scores on the two occasions for some subjects, but not for others; the degree of agreement between measures having reference to two occasions varies from one individual to another, and even the largest rhos imply some considerable disagreement among the rankings of the characteristics at different ages.

The concept of unity among measures of the individual does not seem to allow for the possibility of individual differences in the degree to which such measures are idiomatic within the individual. However, it seems certain that there are individual differences in the extent to which the individual's pattern of adjustment remains constant during the adolescent years.

The data to be analyzed were obtained by means of the U.C. Inventory I, dealing with social and emotional adjustment, which was administered annually to the members of the California Growth Study (*3*). It is a self-report type of inventory, the items of which are organized around eight areas of categories of adjust-

---

[1] Manuscripts dealing with these analyses are in preparation.

[2] Paper read at the C.E.R.A. meeting, March, 1955.

TABLE I.—RELIABILITY COEFFICIENTS FOR CATEGORIES OF ADJUSTMENT

Category	Test-Retest Reliability	Matched-Half Reliability Coefficients		
Name	Occasion	Occasion		
	4	1	2	3
Social maladjustment	.74	.71	.81	.77
Personal inferiority	.68	.62	.60	.45
Overstatement	.71	.64	.58	.52
Family maladjustment	.71	.44	.40	.80
Physical symptoms	.75	.74	.68	.72
Fears	.70	.69	.72	.91
Generalized tensions	.76	.78	.73	.76
School maladjustments	.67	.88	.92	.91

ment. Their designations and reliability coefficients are shown in Table I (7, 8).

Our questions may be formulated as follows:

(1) What is the degree of agreement within an individual among seriatim measures of adjustment obtained during adolescence?

(2) Is the agreement significant for some subjects but not for others?

(3) Is the agreement sufficiently large to permit reasonably accurate predictions from earlier to later ages of the individual?

(4) Is the degree of agreement related to the individual's general adjustment?

THE ANALYSIS

The nature of the questions to be considered and the method of analysis to be employed will be described by means of the data in Tables II and III; the measures are standard scores (based upon grade groups) for each variable at each chronological age. We may determine the extent to which each category of adjustment occupies the same rank among the other categories at each of several ages. Measures of the relationships between the ranks at different ages may be computed by means of Spearman's rank-order method, as for the data in Tables II and III.[3]

The values of rho for the three possible pairings of ages are 0.55,

---

[3] Case 134 in the California Adolescent Growth Study.

TABLE II.—STANDARD SCORES ON THE CATEGORIES OF THE U. C. INVENTORY I BY CHRONOLOGICAL AGE (CASE 134)

CA	Category of Adjustment							
	1	2	3	4	5	6	7	8
11	51	20	50	36	33	72	46	45
13	76	61	57	54	45	63	70	66
17	71	44	49	48	43	72	54	48

TABLE III.—RANK ORDERS OF STANDARD SCORES IN TABLE II

CA	1	2	3	4	5	6	7	8
11	2	8	3	6	7	1	4	5
13	1	5	6	7	8	4	2	3
17	2	7	4	5.5	8	1	3	5.5
Sum	5	20	13	18.5	23	6	9	13.5

0.96, and 0.67, implying a tendency for the variables to occupy similar ranks at different ages. [This type of coefficient of correlation is termed 0-technique by Cattell (*1*)].

The number of rhos to be computed increases rapidly as the number of ages for which data are available increases, so that the amount of computation increases rapidly with an increase in the number of rankings. In any case, a more general question, leading to a more general answer, and involving a limited amount of calculation, might be asked: Is there agreement among the ranks assigned to each category of adjustment for the three ages considered?

Kendall (*4*) proposed his measure of concordance, W, as a suitable technique for answering such a question. If W is found to be significant, we conclude that significant agreement exists among the patternings of the variables at all ages. We may then infer that there is stability in the pattern of an individual's adjustment (although not necessarily in the level of adjustment) as he grows older. The value of W may be computed from data arranged as in Table III, in which the sums of the columns, required for the computation of W, are indicated.

The coefficient of concordance is defined as follows (*2*):

$$W = \frac{\text{Sum of squares between columns}}{\text{Total sum of squares}}$$

$$\text{The between variance} = \frac{(5^2 + 20^2 \cdots 9^2 + 13.5^2)}{3} - \frac{3 \times 8(9)^2}{4}$$

$$= 102.17$$

$$\text{The total variance} = \frac{m(n^3 - n)}{12} = \frac{3(8^3 - 8)}{12} = 126$$

(where m is the number of rows and n the number of columns).

Then, for the data of Table III:

$$\text{The ``between'' variance} = 102.17$$

and the "total" variance = 126

$$\text{so that } W = \frac{102.17}{126} = 0.81$$

The significance of W may be determined by means of the χ^2 test:

$$\chi^2 = \frac{(n - 1)\ (\text{sum of squares between columns})}{(n^3 - n)/12}$$

$$= \frac{7 \times 102.17}{(8^3 - 8)/12} = 17.03,$$

where there are (n-1) degrees of freedom.

For our example, $\chi^2 = 17.03$, which just fails to reach significance at the 0.01 level. We conclude that there is a high probability that the agreement among the patterns of adjustment at different ages is not due to chance.

It might also be of interest to know the mean value of all the rank-order coefficients of correlation in our $m \times n$ table. This mean value may be determined from the following formula (*2*):

$$r^{-1} = \frac{mW - 1}{m - 1} = \frac{3 \times 0.81 - 1}{2} = 0.72$$

In our example, the value of 0.72 is almost identical to the mean value of the three rank-order coefficients already reported for Table III.

TABLE IV.—DISTRIBUTION OF VALUES OF W

Interval	Number of W's
.80–.89	1
.70–.79	1
.60–.69	3
.50–.59	3
.40–.49	6
.30–.39	8
.20–.29	7
.10–.19	1
Total	30

Similar computations were made for thirty boys, all members of the California Adolescent Growth Study. For twenty-eight boys, eight categories of adjustment were available for ranking on each of seven annual occasions (i.e., chronological ages); measures were known for each of six occasions for the other two subjects. The value of W for significance at the 0.01 level is 0.38 for the 7×8 Table, and .44 for the 6×8 Table (W was corrected for continuity and for tied ranks, although these corrections resulted in but minor changes in the value of W).

The distribution of the values of W is shown in Table IV. About 50 per cent of the W's are significant at the .01 level. We conclude that there are individual differences in the presence of significant

TABLE V.—DISTRIBUTION OF VALUES OF THE MEAN OF RANK-ORDER CORRELATIONS

Interval	No. of Mean Rhos
.80–.89	1
.70–.79	1
.60–.69	2
.50–.59	3
.40–.49	2
.30–.39	6
.20–.29	5
.10–.19	7
.00–.09	3
Total	30

agreement among the rankings of the categories of adjustment being considered.

The mean values of all rank-order coefficients of correlation for each subject were also computed according to the formula already indicated. The distribution of mean rhos is shown in Table V. Again, it is obvious that there are individual differences in the degree of agreement among the profiles at different chronological ages. Further, it appears that the prediction of later from earlier profiles must be accompanied by the possibility of large errors. The patterning of the categories of adjustment measured by the U. C. Inventory I does not confirm a hypothesis of unity of growth or of homogeneity of growth for these subjects.

DISCUSSION

It may be that the non-significant values of *W* are the results of changes in the rank orders of the categories in a profile that arose from the unreliability of the measures. If this is so, we might expect that the differences between the standard scores for successive ranks in each profile will be less in the cases of those subjects with the non-significant *W's* than of the other subjects. We determined the distribution of these differences for each of twenty subjects, ten with significant *W's* and ten with non-significant *W's*. The median value for each distribution was then computed; these are shown in Table VI.

It is seen that the medians are very similar for the two groups; that is, the differences between the standard scores for consecutive ranks on any occasion may be as small in the case of the significant as of the non-significant *W's*. The absence of significant agreement among the categories does not appear to be determined solely by unreliability of the measures of adjustment. Some might argue, however, that the presence of several non-significant values of *W*

TABLE VI.—MEDIAN VALUES OF DIFFERENCES BETWEEN STANDARD SCORES REPRESENTED BY CONSECUTIVE RANKINGS FOR TEN SUBJECTS WITH SIGNIFICANT *W*'S AND TEN WITH NON-SIGNIFICANT *W*'S

Group	Median Values									
With significant *W's*	3.0	3.5	3.9	4.0	4.0	4.0	4.1	4.5	5.0	7.0
With non-significant *W's*	3.3	3.8	3.9	4.0	4.1	4.3	4.3	4.5	4.6	5.2

may be a reflection of unreliability of the measures of adjustment (see Table I).

We may now consider the fourth question that was proposed earlier, namely, is general adjustment related to consistency of adjustment over a period of years? The subjects of the California Adolescent Growth Study were rated by each of three psychologists at the end of the school period on each of five categories of adjustment, known as "manifest traits," as follows:

(1) Social prestige
(2) Popularity with same sex
(3) Work adjustments
(4) Heterosexual adjustments
(5) Personality adjustment

The subjects were rated on a seven-point scale, the final rating being the average of the ratings by the three psychologists. Ratings on these five "manifest traits" were available for twenty of our thirty subjects, nine having non-significant *W's* and eleven having significant *W's*. The ratings of each subject on each trait are shown in Figure I; the mean ratings are also indicated for each group. Successful adjustment as rated on the "manifest traits" does not appear to be related to significant agreement among the

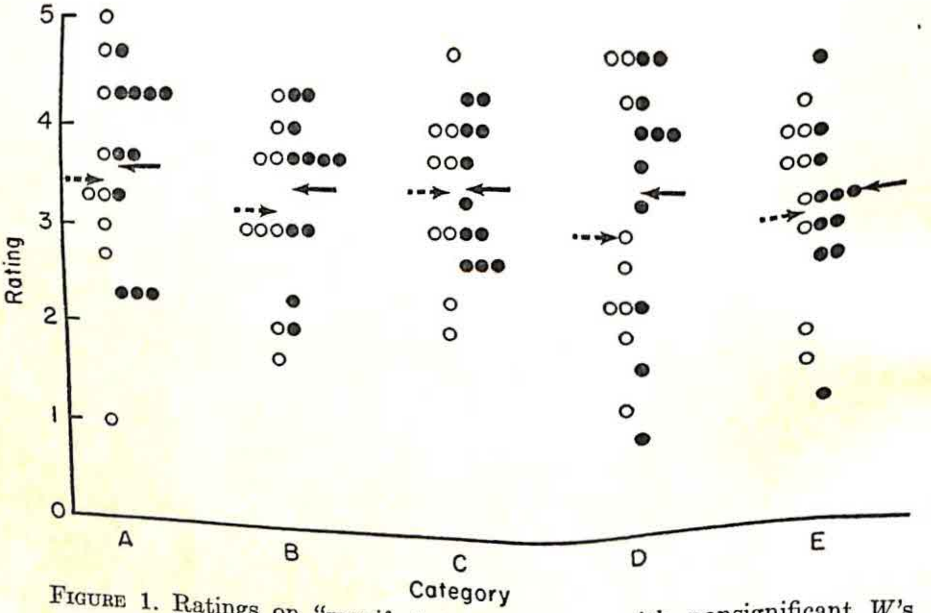

FIGURE 1. Ratings on "manifest traits"; those with nonsignificant *W's* represented by the open circles.

categories of adjustment measured by the U.C. Inventory I during the adolescent years. Stability in the patterning of the individual's self-concept is not related to the psychologists' concept of the subject's adjustment as evaluated by them at the end of the school period. (The rank-order correlations between the value of W and the ratings on each of the five "manifest traits" for the twenty subjects were 0.07, 0.16, 0.09, 0.16, and −0.03.)

SUMMARY

The patterns of adjustment, as revealed by a self-report inventory administered annually to the subjects of the California Adolescent Growth Study, were analyzed by means of Kendall's W to determine the degree of agreement in the patterning of measures on eight categories of adjustment on each of seven annual occasions. The value of W was found to be significant for one half of the subjects. The presence of non-significant W's in some subjects did not seem to arise simply from the unreliability of the measuring instruments since the median differences between the standard scores of consecutive ranks within each profile were about the same for the subjects with significant W's as for those with non-significant W's.

The presence of stability in the patterning of the individual's self-concept on the eight categories of adjustment was not related to successful adjustment as judged by three psychologists at the end of the school period.

It is believed that the nature of the results reported is not favorable to the hypothesis of a high degree of stability in the adolescent's reported self-concept. There are individual differences in the degree of stability; in all cases, prediction of later from earlier measures of adjustment would result in some considerable degree of error.

There remains, of course, the problem of inferring the nature of the individual's behavioral characteristics from a knowledge of the changes in the scores on a self-report inventory. It may be argued that the individual's pattern of adjustment is, in reality, more stable than the profiles imply. Changes in the patterns may reflect changes in expressions of adjustment and maladjustment. An adolescent's relationships within his family may become so unbearable that he comes to the point of indicating good adjustment in the family and at the same time changes from indications of good to poor adjustment in school. Or, again, he may indicate a change from unsatisfactory to satisfactory adjustment in school, but at the same time his "generalized tensions" may become much more prominent in his self-report inventory.

The data analyzed and reported here indicate that the pattern of the adolescent's social and emotional adjustment, as measured by the U.C. Inventory I, may be quite erratic during a seven-year period. A different type of analysis may provide evidence relative to the problem just discussed.

ACKNOWLEDGMENT

The writer is indebted to Dr. Harold E. Jones, who made the original data from the California Adolescent Growth Study available. The statistical analysis, completed by John H. Steinemann, was made possible by a research grant from the Department of Education.

REFERENCES

1) R. B. Cattell, *Factor Analysis: An Introduction and Manual for the Psychologist and Social Scientist,* New York, Harper & Bros., 1952.

2) A. E. Edwards, *Statistical Methods for the Behavioral Sciences,* New York, Rinehart & Co., Inc., 1954.

3) H. E. Jones, "Principles and Methods of the Adolescent Growth Study," *Journal of Consulting Psychology,* 3: 157–159, 1939.

4) M. G. Kendall, *Rank Correlation Methods,* London, Griffin, 1948.

5) F. N. Kerlinger, "The Statistics of the Individual Child: the Use of Analysis of Variance with Child Development Data," *Child Development,* 25: 265–275, 1954.

6) W. C. Olson, *Child Development.* Boston, Heath, 1949.

7) V. S. Taafe, *A Study of the Organization and Consistency of Adjustment Categories.* Unpublished M. A. Thesis, University of California, Berkeley, 1949.

8) C. M. Tryon, *A Study of Social and Emotional Adjustments: the Internal Consistency and Interrelationships of Measurement by Different Techniques,* Ph.D. dissertation, University of California, Berkeley, 1933.

9) F. T. Tyler, "Organismic Growth: P-Technique in the Analysis of Longitudinal Growth Data, *Child Development,* 25: 83–90, 1954.

PUPIL GROWTH IN READING—AN INDEX OF EFFECTIVE TEACHING

HAROLD E. MITZEL and DONALD M. MEDLEY

Division of Teacher Education, Municipal Colleges of New York City

The fact that much of the recent sharp criticism of modern public education has centered on the teaching of reading reflects the importance that the general public attaches to this phase of instruction. The volume of study and experimentation devoted to methods and materials of teaching reading indicates that educators, too, attach great importance to this part of the school curriculum. One important phase of the problem seems to have been neglected, however—the rôle of the teacher in the teaching of reading. A recent review of quantitative studies attempting to evaluate teacher effectiveness reveals few that employ reading growth of pupils as a criterion of effectiveness (*5*). The general conclusion to be drawn from these few studies is that little or nothing is known about what behavioral and personality characteristics distinguish the effective teacher of reading from the ineffective one.

Any attempt to identify such characteristics must begin with the development of a reliable criterion measure of the effectiveness of the teacher on the job to which measures of behavior can be related. This report describes the development of such a criterion.

The chief difficulty that arises when a measure of a teacher's effectiveness is sought in the growth scores of her pupils is the difficulty in separating the teacher's contribution from all of the many other factors which affect the scores. Unless a study is restricted to the teachers in a single school, thereby losing most of its generalizability, different teachers will teach in different schools located in different communities or neighborhoods. And, if elementary-school teachers are being studied, no two teachers will teach the same group of pupils. If the mean growth of one teacher's class in one school is greater than that of another teacher's class in another school, it could be argued that this difference results from differences in schools or neighborhoods, and from differences in pupils, and that perhaps none of the difference can be attributed to a difference in teachers. Suppose one teacher has a high-ability group in a new, well-equipped building in a middle-class neighborhood; another has a "slow" section in a superannuated building in a slum area. Can these teachers be compared on the basis of their pupils' progress?

Allowing for school differences. The problem of school differences can be handled by including in the sample at least two teachers who teach in each school used in the study, and comparing each teacher only with teachers in her school. This has been achieved in the present study by using an analysis of variance with differences between teachers in the same school separated from differences between teachers in different schools. It is possible to test the latter against the former. If they are not greater, school differences may be ignored; if they are greater, comparisons among teachers should be made only within schools.

Controlling class differences. The best way of dealing with class differences among pupils is to assign pupils in the same school to different classes at random. If this is done, then it is safe to conclude that any difference between pupils in different classes in excess of that between pupils in the same class is the result of teacher (or school) differences.

Unfortunately, in this study as in most such studies, the assignment of pupils to classes was not under the control of the investigators. It was necessary to take each teacher with whatever class had already been assigned to her. Naturally, not all the teachers were assigned to the same grade; and within a grade, some teachers had high-ability groups, some low-ability groups, and some even had the local equivalent of a "special" class.

Clearly, there are two important sources of variation among classes which are of primary concern: past achievement and aptitude for learning. It was decided that these two variables would be controlled by the technique of covariance analysis. Each pupil was tested in reading and in mental maturity at the beginning of the school year, and in reading again at the end of the year.

A note on covariance technique. For those unfamiliar with the analysis of covariance, a brief word of explanation may be helpful. The relative amount of reading ability possessed by a pupil is indicated by the deviation of his score on a reading test from the mean of all of the pupils' scores. A reading test administered to a pupil at the end of a school year measures how much reading ability (in comparison with other pupils) he has acquired in his entire school career up to that time. Similarly, a pupil's score on a reading test at the beginning of the year indicates how much he has acquired before coming to his present class. His score on an intelligence test administered at the same time as the initial reading test is a partial measure of his aptitude for learning. The covariance technique predicts, on the basis of performances of all the pupils at the beginning of the year, what score each pupil may be expected to achieve at the end of the year, and then compares the score he actually obtains with that score. It is the pupil whose final score is farthest above his own expected score whom this method regards as having the highest score, instead of the pupil farthest above the group mean.

In other words, the effect of application of covariance adjustments is to start all pupils off as though they had all had the same score on the initial reading test and on the intelligence or mental maturity test.

Other factors. It should be recognized that there may be other differences among these classes (average daily attendance, for instance) that are not measured by the initial tests. Such variables have not been controlled in this study, and remain to contaminate teacher differences, masking to some extent the real differences among teachers. It is necessary to assume that the effect of such variables is negligible.[1] Since, so far as could be ascertained, these variables were not used as a basis for assigning pupils to classes, this assumption is not unreasonable.

COLLECTION OF DATA

Sample. In the fall of 1954, a follow-up study was made to locate those of the one thousand six hundred twenty-eight teachers in the student teaching class of 1953–54 from the municipal colleges who were teaching either as regular teachers or permanent substitutes in New York City public elementary schools. Each teacher who was assigned to teach in grade three, four, five, or six in a school in which there was at least one other such teacher was asked to participate in the part of the project under discussion.

Of the seventy-five teachers who met these criteria, nineteen were eliminated either because the schools were inaccessible, because the principle or one of the teachers in a school did not wish to coöperate, or because one of them failed to accept a proffered teaching position. Participation in the study was voluntary, but partly because of the rather heavy pressure exerted by district superintendents and principals on each teacher to participate, the number of teachers omitted at their own request was small—not greater than three or four. Two principals failed to coöperate; one was the victim of a sudden increase in school population and felt that he had so many problems as a result that he was justified in dropping out; the other maintained that the additional burden on a beginning teacher was too great to justify her even being asked to coöperate.

---

[1] For an analysis of the implications of such assumptions and what it means if they are not fulfilled, see Lindquist (*3*).

Of the fifty-six teachers whose classes were tested in the fall, seven were dropped from the study during the year. Whenever one teacher in a school with two candidates in it withdrew, the other teacher was perforce also lost. Table I shows the distribution of the sample by schools and grades.

Table I.—Grade Levels Taught by 49 Teachers in 19 Schools

School No.	Grade Level Taught				Total
	III	IV	V	VI	
1	1			1	2
2	2	2			4
3	1		1		2
4	1	1	1		3
5	2				2
6	2		2		4
7		1	1	1	3
8	1	1	1		3
9	2				2
10	2				2
11	1	1			2
12	2				2
13			2		2
14		1		1	2
15		1		1	2
16		1	1		2
17	1	2			3
18	4	1			5
19	1	1			2
Totals	23	13	9	4	49

Test data were incomplete at the beginning of the analysis in four classes, so that the covariance analysis to follow was based on forty-five teachers. Three of these forty-five teachers were men. Except for these selective factors, and the requirement that there be two or more new teachers in each school, the sample appears to be representative of beginning teachers in New York City public elementary schools.

Tests used. To measure the reading ability of pupils the authors selected four sub-tests of the California Reading Test (Elementary)—Word Form, Word Recognition, Meaning of Similarities, and Interpretation of Meanings. These four sub-tests have a total of eighty-seven test items. Three sub-tests entitled Meaning of Opposites, Following Directions, and Reference Skills, were omitted in order to shorten the administration time. The California test was chosen as the reading measure for the investigation for the following reasons:

(a) It seems to measure aspects of reading ability consistent with the instructional objectives of the New York City school system.
(b) It offers coverage of a wide range of reading skills with norms given for grade equivalent scores from 2.0 to 9.8.
(c) Equivalent forms are available.
(d) The format does not require complicated marking behaviors on the part of the pupils being tested.

To estimate the learning capacity of the pupils in this study the 1950 Short Form of the California Test of Mental Maturity, Non-Language Section was used. The non-language section, consisting of four sub-tests totaling sixty items, was chosen for the following reasons:

(a) It makes small demand on pupils' reading ability.
(b) It is convenient to administer during a single school period.
(c) It has a range of mental age scores from 5.0 to 20.0.
(d) The format does not require complicated marking behaviors on the part of the pupils being tested.

The manner in which the data were collected implies the following major assumptions:

(a) That the four portions of the *California Reading Test* provide adequate coverage of the reading objectives of New York City teachers.
(b) That the pupils in each class who were present at both ini-

tial and final test administrations are representative of the whole class taught by each teacher.

These assumptions are made in addition to the usual assumptions underlying the statistical techniques applied (*3*).

Administration of tests. During October and November of 1954 the initial reading test and the mental maturity test were administered in fifty-six classes to one thousand three hundred twenty-five pupils. Selected portions of Form AA of the California Reading Test were administered to about one-half of each class and corresponding portions of Forms BB, CC, or DD were administered to the remainder. Form S of the California Test of Mental Maturity, Non-Language Section, was administered after a short rest period. In every instance the tests were given by trained psychometrists who had one or more assistants in the classroom. Pupils marked their responses directly in the test booklets.

During April and May of 1955 the same parts of the equivalent forms of the reading test were administered in forty-nine of the same classes to nine hundred four of the same pupils. The final test, given after a lapse of approximately six months, was administered by the examiner who gave the initial test, and the sequence of the initial testing was preserved as closely as possible.

ANALYSIS OF RESULTS

Differences in final reading score means. An analysis of variance of the final reading scores (that is, number of items passed) of the eight hundred twenty-six pupils in the forty-five classes about which complete data were available is shown in Table II. The mean squares shown there may be interpreted to mean that the difference between two pupils in the same class is, on the average, equal to the square root of one hundred nine or a little more than ten

TABLE II.—ANALYSIS OF VARIANCE OF FINAL READING SCORES OF 826 PUPILS IN 45 CLASSES

Source	D.F.	Mean Square	*F*
Between classes	44	2361.598	21.71**
Within classes	781	108.788	
Total	825		

** Significant at the 0.01 level.

Table III.—Analysis of Variance of Final Reading Scores of 826 Pupils in 45 Classes with Initial Reading Scores Held Constant

Source	D.F.	Mean Square	*F*
Between classes	44	376.384	7.54**
Within classes	780	49.934	
Total	824		

** Significant at the 0.01 level.

points, while the difference between two pupils in *different* classes is, on the average, about forty-nine points.

The fact that the average difference between pupils in different classes is more than four times as great as that between pupils in the same class suggests that there may be significant differences among classes. That there are significant differences is indicated by the *F*-ratio of 21.7, which has a probability of less than 0.01. These differences are probably due in large part to the fact that the classes are not all at the same grade level. One would expect most sixth graders to read better than most third graders. Whether or not these differences may be attributed in part to differences in the quality of the teaching exhibited in the different classes cannot be inferred from this analysis.

Mean differences adjusted for initial reading scores. When the mean squares are "adjusted" to allow for differences in initial scores—when, that is, the departure of each pupil's score from the score predicted for him on the basis of his initial score is used instead of its departure from the mean of all the final scores—the mean square between classes is reduced from two thousand three hundred sixty-two to three hundred seventy-six and the mean square within classes from one hundred nine to fifty. These results are shown in Table III.[2]

The average difference between pupils in the same class is still seven points; indicating that even when initial status is allowed for, there are considerable differences in the effectiveness of a teacher with different pupils. The average difference between pupils in different classes is now about nineteen points. This means that, even if all of the pupils in all of the classes had started out

---

[2] Computational procedures used are given in detail by Johnson (*2*).

TABLE IV.—ANALYSIS OF VARIANCE OF FINAL READING SCORES OF 826 PUPILS IN 45 CLASSES WITH INITIAL READING AND MENTAL MATURITY SCORES HELD CONSTANT

Source	D.F.	Mean Square	F
Between classes	44	386.400	7.97**
Within classes	779	48.496	
Total	823		

** Significant at the 0.01 level.

with the same reading score, the difference between pupils with different teachers would have been greater than that between pupils with the same teacher. The F-ratio of 7.54, significant at the one per cent level, justifies this conclusion.

Adjustment for mental maturity. It is still doubtful whether these differences represent differences in teachers' skill, since the average learning ability (in the sense of mental maturity) of classes might differ. A second "adjustment" was therefore made for mental maturity, with the results shown in Table IV.

Here the average difference between classes increases slightly while the average difference within classes decreases slightly, increasing the F-ratio to 7.97. Most of the differences among pupils in mental maturity seem to have been reflected in differences in initial reading score, and their effect was removed in the first adjustment.

At this stage it seems justifiable to conclude that there are differences in average effectiveness among these forty-five teachers. Allowing for differences among pupils with respect to initial reading scores and mental maturity scores, there remains an average difference among pupils with the same teacher in final reading scores of 6.9 points, while the average difference between pupils with different teachers is 19.7. A given teacher is not equally effective with all pupils, but some teachers are more effective on the average than others.

Estimation of reading growth indices. The next step in the analysis was the estimation of the adjusted mean final reading test score of each class—the reading growth index.[3] Similar indices were computed for the four classes not used in the analysis. For comparison, mean final reading test scores adjusted for initial reading scores only, and unadjusted mean final reading scores were also computed.

---

[3] The method of estimation used is that described by Cochran and Cox (*1*).

Table V shows the class mean final reading scores, unadjusted, adjusted for initial score only, and adjusted for both initial score and mental maturity score. The grade levels and class sizes are also shown. The first, third and fifth columns show the ranks of the classes on each of the three sets of means.

Effects of covariance adjustments on rank order. The rank order of the classes changes considerably with the first adjustment. The class which ranks nineteenth in unadjusted final scores ranks third when adjusted for initial scores; the second highest score moves to the thirty-first place when adjusted. The power of the covariance method is now apparent. In addition to eliminating much variation not related to teacher effectiveness, reducing the average interclass difference from forty-nine to nineteen points, covariance adjustment also rearranges the rank order of the classes. When the initial scores are allowed for, the class which has the second-highest final score is seen to have shown less improvement over initial standing than thirty other classes.

The ranks change relatively little with the second adjustment—that for mental maturity—although the class originally second moves down another three places to the thirty-fourth rank when mental maturity is allowed for. This third set of ranks represents the nearest approximation to the correct rank order of these forty-nine teachers in ability to promote reading growth that can be inferred from these data.

Study of School Differences and Grade Differences. Two more factors which may affect the differences among classes were studied—school differences and grade differences. Table VI shows the results of an analysis of variance of reading growth indices into components between and within schools. The average difference between classes in *different* schools is 5.5 points while the average difference between classes in the *same* school is 4.8; although the former is somewhat larger than the latter, the difference is not significant. It may, therefore, be concluded that differences among schools in the amount of pupils' reading growth were not detected in this study.

Table VII shows an analysis of variance of reading growth indices into components between and within grades. Pupils in dif-

TABLE V.—COMPARISON OF UNADJUSTED, SINGLY ADJUSTED AND DOUBLY ADJUSTED FINAL READING SCORE MEANS OF 49 ELEMENTARY SCHOOL CLASSES

Final Reading Score (Unadjusted)		Final Reading Score Adjusted for Initial Score		Final Reading Score Adjusted for Initial Score and Mental Maturity		Grade Level	Number of Pupils
Class and Rank	Mean	Rank	Mean	Rank	Mean		
1	74.12	2	58.56	2	58.40	5	25
2	69.43	31	47.71	34	47.24	6	21
3	65.52	15	51.10	16	51.14	5	21
4	65.36	6	56.79	6	56.86	5	14
5	65.14	1	60.43	1	60.43	4	7
6	64.60	18	50.67	17	50.80	4	15
7	63.75	24	49.32	26	49.00	4	28
8	63.19	5	56.88	7	56.69	5	16
9	58.70	17	50.95	20	50.50	5	20
10	57.57	10	54.05	12	53.57	3	21
11	57.50	8	55.92	8	56.00	3	24
12	56.65	21	50.12	25	49.29	6	17
13	56.40	19	50.40	19	50.53	4	15
14	56.14	20	50.29	23	49.48	6	21
15	54.77	4	57.23	3	58.08	6	13
16	54.07	30	47.93	29	47.78	4	27
17	53.72	9	54.41	10	54.45	3	29
18	52.95	22	50.05	21	50.05	4	19
19	52.75	3	58.08	4	57.71	5	24
20	52.03	32	47.48	32	47.34	3	29
21	51.66	11	53.91	11	53.97	3	32
22	50.96	7	56.24	5	57.04	5	25
23	50.21	14	52.38	14	52.42	3	24
24	49.92	28	48.31	31	47.62	5	13
25	48.29	29	47.94	27	48.12	3	17
26	47.67	13	53.42	13	53.08	4	12
27	45.56	16	51.06	18	50.75	3	16
28	44.13	27	48.50	28	47.96	4	24
29	43.06	12	53.50	9	54.75	3	16
30	43.00	34	47.32	34½	47.05	3	19
31	42.05	33	47.40	34½	47.05	3	20
32	41.33	39	44.67	42	44.00	4	9
33	40.94	25	49.22	22	49.50	3	18
34	40.86	26	49.00	24	49.43	3	14
35	39.68	23	49.89	15	51.37	3	19

TABLE V.—*Continued*

Final Reading Score (Unadjusted)		Final Reading Score Adjusted for Initial Score		Final Reading Score Adjusted for Initial Score and Mental Maturity		Grade Level	Number of Pupils
Class and Rank	Mean	Rank	Mean	Rank	Mean		
36	39.50	36	46.86	36	46.36	5	14
37	39.29	46	41.59	47	41.47	4	17
38	37.59	35	47.18	30	47.76	3	17
39	35.50	45	42.09	46	41.82	4	22
40	35.11	40	44.47	40	44.79	4	19
41	35.09	41	44.26	41	44.65	3	23
42	33.38	42	44.00	43	43.88	4	8
43	31.13	44	42.12	45	42.25	3	16
44	30.11	43	43.22	38	45.44	3	9
45	30.00	37	44.88	37	45.62	3	8
46	29.86	38	44.82	39	45.41	3	22
47	27.42	47	41.42*	44	42.46	3	24
48	26.85	49	40.08	48	41.38	3	13
49	23.38	48	40.50	49	41.00	3	8
Weighted Grand Mean	48.75		48.75		48.75	Total	904
Range	50.74		20.35		19.43		

* The mean of 41.42 was calculated from a total, a portion of which was derived by regression methods.

ferent grades show a difference on the average of 8.5 points. Whether or not this difference is significantly greater than that between pupils in the same class remains in doubt because the *F*-ratio has a probability between 0.01 and 0.05.

Reliability Estimate. The final step in the covariance analysis is to estimate the reliability of the reading growth indices. The re-

TABLE VI.—ANALYSIS OF VARIANCE OF 49 TWICE-ADJUSTED CLASS MEANS TO TEST FOR SCHOOL DIFFERENCES

Source	D. F.	Mean Square	*F*
Between schools	18	30.461	1.34
Within schools	30	22.778	
Total	48		

TABLE VII.—ANALYSIS OF VARIANCE OF 49 TWICE-ADJUSTED CLASS MEANS TO TEST FOR GRADE EFFECTS

Source	D.F.	Mean Square	F
Between grade levels	3	71.884	3.18*
Within grade levels	45	22.578	
Total	48		

* Significant at the 0.05 level.

liability of a test in any group is the ratio of the variance of the true scores to the variance of the obtained scores of that group. In this instance, the variance of the obtained scores is estimated by the mean square between classes in Table IV. The error variance is estimated by the mean square within classes. If the error variance is subtracted from the obtained score variance an estimate of the true score variance is obtained; the reliability may therefore be estimated to be:

$$r = \frac{386.400 - 48.496}{386.400} = 0.874$$

This coefficient may be interpreted as representing the expected correlation between the mean reading growth indices of the forty-five classes studied and the mean reading growth indices that would be obtained by forty-five other classes taught by the same teachers. Its magnitude indicates that had these teachers been assigned to other, similar classes, their rank order would not have differed greatly from that shown in Table V.

SUMMARY AND DISCUSSION

A follow-up of one thousand six hundred twenty-eight students in New York City municipal colleges located seventy-five of them beginning their teaching careers in grades three, four, five, or six in City public elementary schools which were employing at least two such teachers. Of these teachers, nineteen were eliminated and the classes of the other fifty-six were tested with a reading test and a mental maturity test in the fall of 1954 and forty-nine of them were tested with equivalent forms of the reading test the following spring.

The analysis of variance and covariance was used to estimate the average effectiveness of each teacher in stimulating her pupils to learn to read, taking into account:

(a) Differences in learning aptitude among pupils.
(b) Differences in previous reading achievements among pupils.
(c) Differences in grade level.
(d) Differences in average amounts of improvement in reading ability among pupils in different schools.

The classes of these forty-nine beginning teachers were found to differ widely in the average amount of improvement in reading ability over one year, even when allowance was made for differences in aptitude and amounts learned in previous years. These differences were no greater between classes in different schools than between classes in the same school. In general, there was some evidence that beginning teachers' effectiveness in the area of reading varies for different grades, but the evidence was not conclusive. The reliability of the differences in average reading improvement among classes was such that if the study had been repeated with the same teachers in other classes, the mean growth scores of the classes would have correlated 0.87.

It was therefore concluded that there are substantial differences among beginning teachers in New York City public schools in effectiveness in stimulating pupils to learn to read, and that these differences cannot be attributed wholly to differences among schools nor to differences in pupils' learning ability or previous achievement.

REFERENCES

1) W. G. Cochran, and G. M. Cox, *Experimental Designs*, p. 78, New York, Wiley, 1950.

2) P. O. Johnson, *Statistical Methods in Research*, pp. 246–266, New York, Prentice-Hall, 1949.

3) E. F. Lindquist, *Design and Analysis of Experiments in Psychology and Education*, pp. 323–325 and 328–330, New York, Houghton-Mifflin, 1953.

4) L. J. Lins, "The Prediction of Teaching Efficiency," *Journal of Experimental Education*, *15:* 2–60, 1946.

5) H. E. Mitzel, and C. F. Gross, *A Critical Review of the Development of Pupil Growth Criteria in Studies of Teacher Effectiveness*, Research Series, No. 31, Division of Teacher Education, Board of Higher Education of the City of New York, April 1956.

A FURTHER INVESTIGATION OF THE RELATIONSHIP BETWEEN ANXIETY AND CLASSROOM EXAMINATION PERFORMANCE

ALLEN D. CALVIN, F. J. McGUIGAN, and MAURICE W. SULLIVAN

Hollins College

In a recent paper McKeachie, Pollie, and Speisman (*3*) found that students who were given the opportunity to write comments about their objective examination questions performed better on the last half of their examination than students who were not given an opportunity to comment. They pointed out that this finding "... fitted in with our theory that tension is built up throughout the test and that giving opportunity to comment reduces the increasing tension." (*3*, p. 95). They also reported a relationship between examination scores and McClelland's "need for achievement" (*2*) personality variable. On the basis of their results they concluded: "Anxiety inhibits performance. Giving students an oportunity to write comments aids not only in reducing the threat but also in channeling the release of anxiety" (*3*, p. 98).

Since McKeachie et al., themselves, noted after their first experiment that "our results seemed too good to be true", an attempt to replicate their findings seems called for. In addition, the fact that they had no direct measure of anxiety indicates the need for an experiment where such a measure is provided. The following study is therefore an attempt to replicate McKeachie et al.'s results, and to test their "anxiety reduction" hypothesis.

METHOD

Subjects. The *S*s, one hundred and fifty-two undergraduate female students from Hollins College, were taken from five classes as follows: sixty-one and thirty-five from two introductory psychology classes, seventeen and twenty-one from two introductory Spanish classes, and eighteen from a Spanish literature class. In each case the entire class participated.

Procedure. In an attempt to distinguish between high anxious and low anxious *S*s, the A-Scale with the biographical inventory as described by Taylor (*6*) was given. The Otis Higher Examination of Mental Abilities (*5*) was also given. Both tests were administered at the beginning of each course. The rest of the procedure was similar to that used by McKeachie et al., who modified their methodology slightly from group to group. The specific procedure used in our experiment was as follows: The examination consisted of randomly presented, multiple choice items and was the first examination given in the course for each group. Each class was divided at random into an experimental and a control group. The experimental group in each class received answer sheets with the instructions: "Put an X through the best answer for each item. Feel free to make any comments about the items in the space provided," and the control group received answer sheets with the instructions: "Put an X through the best answer for each item. Do not mark in the space to the right of your answers."

RESULTS[1]

In all five classes, the experimental group made fewer errors on the last half of the examination than did the control group. Tests between experimental and control groups in each class, however, failed to reach the 5% point of significance. An F test between classes was made and no significant difference obtained, so the findings for all the classes were combined and the one-tailed binomial expansion indicated that the present results could be expected by chance approximately three times out of one hundred. Accordingly, the null hypothesis that there was no difference in performance between experimental and control groups on the last half of the examination was rejected.

The largest introductory psychology class permitted us to make an analysis in terms of the A-Scale. The sixty-one *S*s were divided at the median and classified as low and high anxiety groups. The high anxiety *S*s in the experimental group were significantly worse ($p < 0.04$, one-tailed *t* test) on the first half of the examination, but did not differ significantly from the low anxiety *S*s on the second half (See Table I). An evaluation, in terms of gain from the first

---

[1] Data gathered in connection with another study made it possible to test McKeachie et al.'s "anxiety reduction" hypothesis on a limited number of *S*s using another measure of "anxiety" [the Palmer perspiration index described by Mowrer, (*4*)]. This analysis yielded non-significant results.

TABLE I.—ERRORS ON FIRST AND LAST HALVES OF EXAMINATION

Group	First Half	Second Half
Experimental		
High anxiety	9.58	7.25
Low anxiety	7.24	7.35
Control		
High anxiety	9.00	8.82
Low anxiety	8.43	6.78

half of the examination to the second, showed that the high anxious experimental *S*s were significantly superior to the low anxious experimental *S*s ($p < 0.005$). When a covariance to remove the effect of initial scores on gain scores was run the resulting F was still significant ($p < 0.02$). This difference in gain scores is indicated in Table I by the sharp reduction in errors for the high anxiety experimental group on the second half of the test, and by the slight increase in errors for the low anxiety experimental group. For the control group, the high anxiety *S*s made more errors than the low anxiety *S*s (although not significantly more) on *both* the first and second half of the test; while the high anxiety *S*s did improve somewhat on the second half of the test the low anxiety *S*s improved even more.[2]

The correlation between total number of comments and number of errors on the second half of the examination was −0.19 for the high anxiety experimental group, and −0.40 for the low anxious experimental group, neither of which was significant.

The correlation between Otis scores and anxiety scores was −0.05.

DISCUSSION

The finding that all five experimental groups were superior to their corresponding controls corroborates the findings of Mc-

---

[2] In a personal communication from W. F. Smith, we have learned that he and F. C. Rockett, did a related study using the Sarason anxiety scale. They did not analyze their results to see if high anxiety *S*s showed a greater gain than low anxiety *S*s on the second half of the examination, but they varied type of instruction and although the effect of type of instruction was not significant, they found a significant interaction between type of instruction and anxiety level. High anxiety *S*s benefited more from anxiety reducing instruction.

Keachie et al. The failure of the individual intra-class comparisons to reach acceptable levels of statistical significance may well be due to the smaller *N*s in our classes than in McKeachie et al.

The sharp drop in errors for our high anxiety experimental group on the second half of the examination is similar to the behavior of McKeachie et al.'s low need for achievement group and supports their anxiety reduction hypothesis.

McKeachie et al. report a correlation for their low need for achievement group of 0.73 between scores on the last half of the examination and number of comments, and for their high need for achievement group of 0.05. Our correlations, between number of errors on the last half of the examination and number of comments, of course, reverses the sign so that in both McKeachie et al.'s *S*s, and ours, those *S*s who made the most comments performed the best. Neither of our correlations, however, approached the magnitude of McKeachie et al.'s low need achievement group, and our low A-Scale group actually had a higher correlation than our high A-Scale group, although neither of our correlations was significant. Rather than speculate on these results at this time, it seems wiser simply to note that there is a trend for those *S*s who make more comments to make fewer errors on the last half of their examinations.

The practically zero correlation between Otis scores and anxiety scores certainly does not point toward any relationship between anxiety and intelligence in our sample, although we cannot of course definitely rule out the possibility that there may exist differences in intellectual makeup between our high and low anxiety *S*s which are not shown by the Otis (*1*, *7*).

The present results, taken together with those of McKeachie et al., indicate the applicability of "drive reduction learning theory" to the classroom situation.

SUMMARY

In an effort to confirm findings by McKeachie et al., five undergraduate classes were divided into experimental and control groups. The experimental groups were allowed to comment on items in an objective examination while the control groups were not. When the results of the five classes are combined, the experimental *S*s performed significantly better than the controls on the second half of the examination. The largest class with an N of sixty-one was

divided into high and low anxious *S*s on the basis of their A-Scale scores. Anxiety was found to be a significant variable in terms of improvement from the first half of the test to the second in the experimental group. It was also found that those *S*s who made the most comments showed the greatest improvement, but this relationship was not significant. Implications of these findings are discussed.

REFERENCES

1) A. D. Calvin, P. B. Koons, Jr., J. L. Bingham, and H. H. Fink, "A Further Investigation of the Relationship Between Manifest Anxiety and Intelligence," *Journal of Consulting Psychology*, 19: 280–282, 1955.

2) D. C. McClelland, J. W. Atkinson, R. A. Clark, and E. L. Lowell, *The Achievement Motive*, New York, Appleton-Century-Crofts, 1953.

3) W. J. McKeachie, D. Pollie, and J. Speisman, "Relieving anxiety in classroom examinations," *Journal of Abnormal and Social Psychology*, 50: 93–98, 1955.

4) O. H. Mowrer, *Psychotherapy*, New York, Ronald Press, 1953.

5) A. S. Otis, *Otis Self-Administering Tests of Mental Ability*, Chicago, World Book Company, 1950.

6) Janet A. Taylor, "A Personality Scale of Manifest Anxiety," *Journal of Abnormal and Social Psychology*, 48: 285–290, 1953.

7) Janet A. Taylor, "The Taylor Manifest Anxiety Scale and Intelligence." *Journal of Abnormal and Social Psychology*, 52: 347, 1955.

SEARCHING ORIENTATION AND CONCEPT LEARNING[1]

GABRIEL M. DELLA-PIANA

Public School District No. 111, Highwood, Illinois

Results of research during the first three decades of the twentieth century established that learning may be enhanced by giving the learner "knowledge of results." McGeoch and Irion (*8*, pp. 265–268), in summarizing much of this research, conclude that knowledge of results has its greatest learning value when obtained with optimum immediacy, specificity, and frequency. Theoretical interpretations of these conclusions (*1, 2, 9, 11*), suggest three processes by which knowledge of results may influence learning: (a) showing progress and thus "motivating" the learner. (b) presenting a standard and thus "guiding" the learner's trial responses. (c) indicating errors and allowing the learner to find out why the response was wrong, thereby "activating a searching orientation" in the learner.

The present study assumes that a "searching orientation" may be induced by manipulating the method by which the subject obtains "informative feedback." On the basis of this assumption, predictions are made concerning how variations in amount of searching behavior may influence the learning of concepts. "Informative feedback procedures" is a brief term used in this paper in lieu of the longer phrase "procedures for giving a learner knowledge of which responses are incorrect." "Searching orientation" refers to the subject's hypothesizing as to why a response was not correct. The hypotheses thus developed by the subject are tested by him each time he receives knowledge of results. Presumably, this process leads to the learning, not only of correct responses, but also of reasons underlying the correct responses. Thus, the searching orientation was presumed to act as an intervening variable through which knowledge of results might influence learning. Underwood (*11*, pp. 414–417) mentions this "searching orientation" (he calls it "attempt to discover") as a variable probably accounting for the learning value of knowledge of results. Jones (*6*) and McConnell (*7*) conducted studies of classroom teaching procedures in which a "searching orientation" seemed to be induced by feedback procedures and appeared to facilitate "meaningful" learning.

---

[1] A more complete presentation of the hypotheses, data, and procedures of this study is available in G. M. Della-Piana's "Two Experimental Feedback Procedures: A Comparison of their Effects on the Learning of Concepts," unpublished Ph.D. dissertation, 1956, on file in the University of Illinois Library. Also available as microfilm, Publication No. 16,386, University Microfilms, 313 No. First St., Ann Arbor, Michigan, $1.10. The study was done under the direction of Professors Lee J. Cronbach and J. T. Hastings.

METHOD

Two experimental treatments. Two informative feedback procedures were developed to produce variations in "searching orientation" of subjects. These procedures constituted our two experimental treatments: one (dependency treatment) in which a subject was immediately told the correct answer when he gave an incorrect response, and the other (searching treatment) in which a subject was told to keep trying until he "discovered" the correct answer. Our searching treatment is actually the standard "correction procedure" used in experimental learning studies (*10*, p. 13). In the present study it is called the "searching treatment" because it was assumed to induce the subject to develop hypotheses as to why his responses were wrong. The dependency treatment was developed to produce a lesser degree of searching behavior than that produced by the searching treatment. We attempted to attain this aim by varying the feedback procedure so as to force the subject to make a response antagonistic to searching. The response we attempted to induce by the dependency treatment was the relatively non-inquiring one of "looking to the experimenter for the answers." Differences between treatments, other than the theoretically crucial difference, were minimized. For example, "time on item" differences were minimized; however, the degree of "searching" engaged in during that time was varied. We did not directly investigate the psychological process of "searching." Instead, we investigated hypothesized differential effects which variations in "degree of searching" presumably had on the learning of concepts.

Two ability levels. An attempt was made to utilize a two-by-two treatments by levels design. Guilford's Figure Exclusion Test (*3*) was used as a basis for dividing subjects into ability levels since this test measured factors related to those called for by our learning task. Of the eighty-two students tested, we selected the forty extremes on the Guilford test (twenty in each treatment). Unfortunately, almost half of the differences between ability level groups on the Figure Exclusion Test were lower than two times our standard error of the difference between scores. Thus, we had very little confidence in the differences between ability level groups on this test.

No statistically significant differences between ability level groups were obtained in this study on any of the dependent variable measures. We may say that differences between levels, if they existed, were not strong enough to show up in our sample under the conditions noted above. Therefore, further data on "levels" will not be reported in this paper.

The concept learning task. An inductive concept formation task developed by Heidbreder and Ivy (*4*) was used. We chose this task for two reasons: First, it may be learned either through "rote association" (naming the concepts without being able to identify or define their distinguishing characteristics) or through more "meaningful association" in which the learner can name and define the concept (*4*, p. 124). This was an essential characteristic of the task used in this study since we expected our two feedback procedures to differ with respect to the extent to which they facilitated one or another of these two kinds of learning. Secondly, the task is similar in many respects to certain school learning tasks where drill procedures similar to our experimental feedback procedures are utilized; e.g., in the learning of basic arithmetic facts or sight vocabulary.

As used in this study, "concepts" refer to certain common attributes of a group of geometrical designs presented on cards. The particular concepts we utilized were two of color (e.g., "blueness"), two of shape (e.g., "diamond-shapedness"), and two of number (e.g., "fourness"). Each of the six concepts was given a nonsense-syllable name, such as BLAG or LORB.

The concepts were presented to the subject embedded in drawings on 3″ by 5″ cards. The subject was presented the drawings in ten separate series of six drawings each. Each of the six drawings in a series contained one of the six concepts. Each concept was represented in each series; however, in addition to the constant component of color, shape, or number, which defined the concept, it had a variable component from series to series. Thus, the concept "fourness" was always represented by a drawing of four geometric designs. However, in one series it was "four, pink, triangular designs"; in another series it was "four, orange, rectangular designs"; and so on; but it was always four designs.

The subject was asked to learn the nonsense syllable names of designs presented to him singly and successively in series after series of cards. The conditions made it possible for the subject to discover that a given nonsense syllable, though used as the name of very different designs in the different series, was always applied to designs which, taken together, possessed a common and distinguishing characteristic. Thus, "BLAG" was always used as the name of a design of the concept "fourness" (the design having four objects) though the shape and color of the design varied from series to series.

DATA AND SCORES

The hypotheses of the study will have more meaning to the reader after a brief description of the data and scores has been digested. There are three sets of scores: learning series scores, post-treatment scores, and scores combining learning series and post-treatment measures.

Learning series scores. In the learning series a concept was said to be "attained" when a subject named it correctly on the first trial (and all subsequent trials) of at least two consecutive series and when he never named it incorrectly after these two consecutive series. Thus, "concept attainment" refers to ability to give nonsense syllable names of concepts without regard to whether one knows the distinguishing feature of the concept.

The raw data of the learning series consists of an error tally on a concept for each trial (within a series) in which the subject names the concept incorrectly. From these data we obtained a sum of series required for concept attainment and a sum of number of concepts attained.

Two means, derived from these data, were used in testing the hypotheses of this study.

(a) Mean number of concepts attained (NCA). This was obtained by summing the total number of concepts attained by each person in a treatment and dividing this sum by the total number of persons in that treatment.

(b) Mean number of series to attain concept (SCA). This was obtained by summing the total number of series required for attainment of a given concept by each person in a treatment and dividing this sum by the total number of persons in that treatment.

Post-treatment scores. Following the learning series, each subject was given two tests: first, an open-ended definitions test and then a multiple choice definitions test. These tests presented the subject with a list of the nonsense syllable names of the concepts and required the subject first to write a definition of the distinguishing feature of the concept and then to select, from among several alternatives, the distinguishing feature of the concept. The test score was the number of concepts correctly defined. On the open-ended test this was called a "recall of meaning score" and on the multiple choice test it was called a "recognition of meaning" score.

Two means, derived from these scores, were used in testing the hypotheses of this study.

(a) Mean recall of meaning (RcM). This was obtained by summing the total number of meanings recalled by each person in a treatment and dividing this sum by the total number of persons in that treatment.

(b) Mean recognition of meaning (RgM). This was obtained by summing the total number of meanings recognized by each person in a treatment and dividing this sum by the total number of persons in that treatment.

Thus, the post-treatment measures allowed us to determine the mean number of concept definitions recalled and recognized by each treatment group.

Combined learning series and post-treatment score. One mean was obtained by use of combined measures.

(a) Mean meaningful-concept attainment (MCA). This was obtained by summing the percentage (decimal fraction) of attained concepts for which each person in a treatment could recall the correct meaning and dividing this sum by the total number of persons in that treatment.

Thus, combining learning series and post-treatment measures allowed us to determine the mean percentage of "attained" concepts for which a treatment group could recall the definitions.

TABLE I.—SUMMARY OF TESTS OF SIGNIFICANCE OF DIFFERENCES BETWEEN TREATMENT MEANS ON FOUR VARIABLES
(d.f. = 1 and 36) (Treatment N = 20)

	Dependent Variable Measures			
	Mean of concepts named	Mean of concept definitions recalled	Mean of concept definitions recognized	Mean fraction of named concepts recalled
Searching treatment	4.40	4.75	5.10	.97
Dependency treatment	3.75	3.60	4.05	.81
Difference	.65	1.15	1.05	.16
Mean square*	4.22	13.22	11.03	.24
F ratio†	1.84	8.58‡	9.59‡	12‡

* Between treatments.

† Using within treatments mean square.

‡ Significant beyond 0.01 level.

TESTS OF HYPOTHESES

Results of the major hypotheses of the study are presented in Table I. Although the study was set up with a two-treatments-by-two-levels design no data on levels or treatment-levels interactions are presented here because of non-significance of "levels" differences on both ability and dependent variable measures.

Hypothesis I. The mean number of concepts attained (NCA) will be greater for subjects in the "searching" treatment than it will be for subjects in the "dependency" treatment.

As Table I shows, treatment differences were in the predicted direction, but were not statistically significant. Incidentally, they were significiant for low ability subjects taken alone.

Hypothesis II. The mean number of series required for concept attainment (SCA) will be less for subjects in the "searching" treatment than for subjects in the "dependency" treatment.

A sign test for the significance of the direction of differences showed that only seven out of twelve differences between treatments were in the predicted direction. This gives us a z-score[2] of only 0.058, definitely not statistically significant.

Hypothesis III. The mean number of concepts for which one can recall the meaning (RcM) will be greater for "searching" treatment subjects than for "dependency" treatment subjects.

Hypothesis IV. The mean number of concepts for which one can recognize the meaning (RgM) will be greater for "searching" treatment subjects than for "dependency" treatment subjects.

---

[2] (raw score − mean)/standard deviation.

Hypothesis V. Of all the concepts attained by subjects, the mean percentage of these concepts for which the subjects can recall the meaning (definition) will be greater for "searching" treatment subjects than for "dependency" treatment subjects.

As Table I indicates, concerning Hypotheses III, IV, and V, in all cases the overall treatment differences are statistically significant in the hypothesized direction.

DISCUSSION

Our interpretation of the above results may be summed up as follows: Within the restrictions of the procedures of this study, it is shown that subjects who are allowed to search for a correct answer (S-group) compare with subjects who are told the correct answer (D-group) in the following ways:

(a) No significant differences in number of concepts named correctly.

(b) No significant differences in number of series presentations required to name concepts correctly.

(c) S-group learns to recall and recognize significantly more definitions of concepts than D-group.

(d) S-group (more than D-group) learns to recall definitions of a greater percentage of concepts they learned to name correctly.

There are certain qualifications and suggestions for further study which may be noted at this point.

A very important suggestion arises as a consequence of the positive results of Hypotheses III, IV, and V. The rationale for these hypotheses was that the "searching" treatment, more than the "dependency" treatment, induced a set for "formulating and testing hypotheses" and that this hypothesizing and testing activity was the main factor in accounting for treatment differences in meaningfulness of learning. Now that significant differences have been found, it is advisable to investigate further along two lines suggested below.

First, an attempt should be made to get at the differences in problem solving processes of subjects in each of the treatments. Secondly, an attempt should be made to get information on differences in incentive factors between treatments. That is, now that we have found genuine differences between treatments on our dependent variable measures, we want to know more specifically how these treatments influence learning. Thus, the question remains open and testable: To what extent are observed differences between treatments due to "hypothesizing and testing activity" and to what extent are they due to "differences in motivational characteristics"?

Further research on Hypotheses I and II is also suggested by two observations. First, an association interference variable was not equated between treatments and operated against the prediction. Only one incorrect association could be made by dependency treatment subjects before they were given the correct name of a concept card. It was possible for "searching" treatment subjects to make as many as five incorrect associations with a given card. For experimental purposes, it would be possible and desirable to equate this variable between treatments. Secondly, limiting ourselves to only ten series of cards seriously restricted the range of concept attainment scores for our sample of subjects. What effect more trials, and greater discrimination, would have on our results is a matter for further experimentation.

SUMMARY

Two informative feedback procedures were developed to produce hypothesized variations in searching orientation of subjects. Searching orientation was not directly observed; however, predictions were made concerning how variations in searching orientation would effect performance in naming and defining concepts. As predicted, the group assumed to do more searching learned to give significantly more definitions of concepts. This major positive result suggests the value of more direct study of the hypothesized psychological processes presumably induced by the feedback procedures. The hypotheses that the searching group would learn to name more concepts in fewer trials were not supported.

REFERENCES

1) R. B. Ammons, "Knowledge of Performance: Survey of Literature, Some Possible Applications, and Suggested Experimentation," WADC Technical Report 54-14. Wright-Patterson Air Force Base, Ohio, February 1954.

2) W. F. Book and L. Norvell, "The Will to Learn: An Experimental Study of Incentives in Learning," *Pedagogical Seminary*, 29: 305–362, 1922.

3) J. P. Guilford, R. F. Green, and P. R. Christensen, "A Factor-Analytic Study of Reasoning Abilities. II. Administration of Tests and Analysis of Results, *Report of Psychology Laboratory,* No. 3. Univ. Southern Calif., 1951.

4) E. Heidbreder, "The attainment of concepts: III. The Process," *Journal of Psychology,* 24: 93–138, 1947.

5) M. L. Ivy. "The Production of Conceptions of Spatial Forms, Colors, and Numbers," Master's Thesis. Wellesley College, 1945.

6) R. S. Jones, "Integration of Instructional with Self-Scoring Measuring Procedures," Doctoral dissertation. Ohio State Univ., 1951.

7) T. R. McConnell, "Discovery Versus Authoritative Identification in the Learning of Children," *University of Iowa Studies in Education.* 9: No. 5, 13–62, 1934.

8) J. A. McGeoch and A. L. Irion, *Psychology of Human Learning,* New York, Longmans Green, 1952.

9) H. Seashore and A. Bavelas, "The Functioning of Knowledge of Results in Thorndike's Line Drawing Experiment," *Psychological Reviews,* 48: 155–164, 1941.

10) K. W. Spence, "The Solution of Multiple Choice Problems by Chimpanzees," *Comparative Psychology Monographs,* 1939. 15, (3).

11) B. J. Underwood, *Experimental Psychology.* New York, Appleton-Century-Crofts, 1949.

PUBLICATIONS RECEIVED

Ruth Barry, Harvey Hall, Esther Lloyd-Jones, Thomas Shrewsbury, Jean Wellington and Beverly Wolf. *Case Studies in College Student-Staff Relationships.* New York: Bureau of Publications, Teachers College, Columbia University, 1956, pp. 115. $2.00.

Glenn Myers Blair. *Diagnostic and Remedial Teaching.* New York: The Macmillan Company, 1956, pp. 409. $5.00.

Paul H. Bowman, Robert F. DeHaan, John K. Kough and Gordon P. Liddle. *Mobilizing Community Resources for Youth.* Chicago: University of Chicago Press, 1956, pp. 138 (paper). $2.50.

Jerome S. Bruner, Jacqueline J. Goodnow and George A. Austin. *A Study of Thinking.* New York: John Wiley & Sons, Inc., 1956, pp. 330. $5.50.

Elizabeth Cameron. *Reader and Workbook: All About Baby.* (Pre-Primer Level.) New York: Wonder Books, Inc., 1956, pp. 60 (paper) $0.25.

Elizabeth Cameron. *Reader and Workbook: Away I Go.* (Pre-Primer Level.) New York: Wonder Books, Inc., 1956, pp. 60 (paper). $0.25.

Elizabeth Cameron. *Reader and Workbook: A Wish for Billy.* (First Reader Level.) New York: Wonder Books, Inc., 1956, pp. 60 (paper). $0.25.

Elizabeth Cameron. *Reader and Workbook: We Live on a Farm.* (Primer Level.) New York: Wonder Books, Inc., 1956, pp. 60 (paper). $0.25.

A Description of the College Board Achievement Tests. New York: College Entrance Examination Board, 1956, pp. 60 (paper). $0.50.

A Description of the College Board Scholastic Aptitude Test. New York: College Entrance Examination Board, 1956, pp. 64 (paper). $0.50.

The Exceptional Child Faces Adulthood. Proceedings of the 1955 Spring Conference. Langhorne, Pennsylvania: The Woods Schools, 1956, pp. 114 (paper). $1.00.

Estelle J. Foote. *Six Children.* Springfield, Illinois: Charles C Thomas, 1956, pp. 317. $5.00.

ERIC F. GARDNER and GEORGE G. THOMPSON. *Social Relations and Morale in Small Groups.* New York: Appleton-Century-Crofts, Inc., 1956, pp. 312. $6.00.

J. W. GIACHINO and HENRY J. BEUKEMA. *American Technical Society's Drafting.* Chicago: American Technical Society, 1954, pp. 243. $3.80.

WILLIAM S. GRAY. *The Teaching of Reading and Writing.* Paris: UNESCO, 1956, pp. 286 (paper). $3.00.

LOWRY W. HARDING. *Essays in Educology.* Dubuque, Iowa: William C. Brown Company, 1956, pp. 186 (paper). $2.50.

LESLIE HOUSDEN. *The Prevention of Cruelty to Children.* New York: Philosophical Library, Inc., 1956, pp. 406. $7.50.

EDWARD S. JONES. *College Graduates and Their Later Success.* Buffalo, New York: University of Buffalo, 1956, pp. 121–208 (paper). $1.00.

D. B. KLEIN. *Mental Hygiene: A Survey of Personality Disorders and Mental Health.* New York: Henry Holt and Company, 1956, pp. 654. $6.75.

HAL LEHRMAN. *Portrait of Israel—Myth and Reality.* New York: Public Affairs Pamphlets, 1956, pp. 28 (paper). $0.25.

HERTA LOEWY. *Training the Backward Child.* New York: Philosophical Library, Inc., 1956, pp. 166. $3.75.

HERSCHEL T. MANUEL. *Taking a Test: How to Do Your Best.* Yonkers-on-Hudson, New York: World Book Co., 1956, pp. 77 (paper). $0.96.

MARGARET MCEATHRON. *Your Child Can Learn to Read.* New York: Grosset & Dunlap, 1956, pp. 92. $1.50.

LYLE L. MILLER. *Increasing Reading Efficiency.* New York: Henry Holt and Company, 1956, pp. 303 (paper). $2.95.

THERESA G. MULLER. *The Foundations of Human Behavior: Dynamic Psychology in Nursing.* New York: G. P. Putnam's Sons, 1956, pp. 254. $4.50.

OTTO POLLAK. *Integrating Sociological and Psychoanalytic Concepts: An Exploration in Child Psychotherapy.* New York: Russell Sage Foundation, 1956, pp. 284. $4.00.

FREDERICK B. REA. *Alcoholism: Its Psychology and Cure.* New York: Philosophical Library, pp. 143. $3.50.

JEAN ROLIN. *Police Drugs.* Translated by LAURENCE J. BENDIT. New York: Philosophical Library, Inc., 1956, pp. 194. $4.75.

JEAN-PAUL SARTRE. *Being and Nothingness: An Essay on Phenomenological Ontology*. Translated by HAZEL E. BARNES. New York: Philosophical Library, Inc., 1956, pp. 638. $10.00.

MARGUERITE SECHEHAYE. *A New Psychotherapy in Schizophrenia.* Translated by GRACE RUBIN-RABSON. New York: Grune & Stratton, Inc., 1956, pp. 199. $4.50.

LEONARD SMALL. *Rorschach Location and Scoring Manual.* New York: Grune & Stratton, 1956, pp. 214. $6.50.

ROSS STAGNER. *The Psychology of Industrial Conflict.* New York: John Wiley & Sons, Inc., 1956, pp. 550. $8.00.

GEORGE G. STERN, MORRIS I. STEIN and BENJAMIN S. BLOOM. *Methods in Personality Assessment: Human Behavior in Complex Social Situations.* Glencoe, Illinois: The Free Press, 1956, pp. 271. $6.00.

ANSELM STRAUSS. *The Social Psychology of George Herbert Mead.* Chicago: The University of Chicago Press, 1956, pp. 296 (paper) $1.50.

A. H. TURNEY. *Behavioral Potentials and Kinetics: A Phenomenological Approach to the Study of Attitudes and Other Behavior Determinants.* Lawrence, Kansas: University of Kansas School of Education, 1956, pp. 28 (paper).

GEORGE M. WHEATLEY and GRACE T. HALLOCK. *Health Observation of School Children.* Second Edition. New York: McGraw-Hill Book Company, Inc., 1956, pp. 488. $6.50.

WERNER WOLFF. *Essentials of Psychology.* New York: Grune & Stratton, Inc., 1956, pp. 385. $6.50.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 48 May, 1957 Number 5

TEACHERS, PEERS AND TESTS AS PREDICTORS OF ADJUSTMENT

CHARLES A. ULLMANN

U. S. Civil Service Commission*

In an earlier study (*1*), it was suggested that "ratings appear to be better predictors of that aspect of maladjustment which had to do with society's reaction to acted-out behavior, and self-descriptive data appear to be better predictors of what the person himself will chose to do." This report examines the implications of this hypothesis by following the subsequent course of events for three hundred and thirty-one children who participated in the previous study and provides validity data for the predictor measures.

A separate criterion was needed for the examination of each aspect of the hypothesis, i.e., one for events involving actions initiated by a representative of a social institution, such as the school, in giving an award or the police in making an arrest; and one for events in which the individual person initiates action voluntarily, such as in deciding to obtain mental health assistance or to withdraw from school.

The particular criterion measures available for studying the stated hypothesis were: (1) designation by the school as an honor student in the mid-eleventh grade, and (2) withdrawal before graduation from the twelfth grade.

The population on which the first part of this report is based includes all the students in eleven ninth-grade classes of three Prince Georges County (Maryland) junior high schools which participated in the previous study in 1950. This report also includes an item analysis based on one of the tests given to fifty-

---

* Formerly Psychologist, Prince Georges County (Md.) Mental Health Study Center, a demonstration activity of the U. S. Public Health Service.

Bureau Ednl. Re
Dated

one ninth-grade boys described by teachers as "seriously maladjusted" (out of a population of four hundred and four boys in the nine junior high schols included in the previous study).

I

The predictor variables for which scores were obtained at the time pupils were in the ninth grade were:

(1) Forced Choice Test—devised to obtain teacher ratings on pupils
(2) California Test of Personality—"Self" score
(3) California Test of Personality—"Social" score
(4) SRA Youth Inventory—Basic Difficulty score
(5) Sociometric rating

The criterion variables in this study are:

(1) Appearance of name on high school honor roll in February, 1953, that is, at mid-eleventh grade
(2) Withdrawal from school prior to graduation from twelfth grade, that is, prior to June, 1954.

The use of honor roll as a criterion had certain inherent limitations. A small number of children did not have an opportunity to be considered for honor roll status either because of entry into a high school which did not make such awards or because they had moved from the community under conditions which were presumed to indicate continuance in school but which did not allow information on honor-status to be obtained readily. Designation as an honor student is, furthermore, subject to the operation of numerous social influences, difficult to define and assess, and possibly worthy of serious sociological study in their own right. The correlations reported below, therefore, may be somewhat attenuated because of the inaccuracy of the criterion.

The data on withdrawal from school prior to graduation were more satisfactory. Of the original group of three hundred and thirty-one students, information on in-school or out-of-school status was obtained on all but fifteen children (seven boys, eight girls), or 4.5 per cent of the three hundred and thirty-one students (one hundred and sixty-nine boys, one hundred and sixty-two girls) in the population which was followed.

Comparison of scores made by each criterion group on all the five predictor variables is shown in Table I. The means of the honor roll group and the means of the graduates are nearer the "ad-

TABLE I.—SCORES OF HONOR VS. NON-HONOR AND GRADUATE VS. NON-GRADUATE CHILDREN

	a	b	c	d	e	f	g	h	i	j	k	l
	Honor roll	Not on honor roll	Diff	t	r*	Graduate	Non-graduate	Diff	t	r*	r†	r‡
N	39 Score	277 Score				255 Score	61 Score					
Forced choice	25.03	19.82	5.21	6.9	.45	21.42	16.46	4.96	6.5	.46	.46	.27
Sociometric	35.08	51.49	16.41	5.9	.45	46.89	60.23	13.34	4.9	.39	.42	.22
California self	73.95	68.92	5.03	3.0	.22	70.54	65.38	5.16	2.7	.25	.24	.10
California social	75.92	70.90	5.02	3.2	.24	72.49	67.48	5.01	2.9	.26	.25	.20
SRA basic difficulty	12.00	14.96	2.96	1.8	.13	14.15	16.48	2.33	1.1	.11	.12	.14

* Biserial r.

† Triserial r.

‡ Point-biserial r for sex.

justed" pole of each of the variables than are the means for the non-honor roll group and non-graduates. (Signs of correlation coefficients are accordingly reflected wherever appropriate throughout this paper, i.e., all coefficients are shown as positive.) For each variable except the SRA Basic Difficulty, the differences between means exceed the one per cent level of confidence, that is, exceed a t-value of 2.58, and may therefore be regarded as statistically stable.

The relationships between each predictor variable and both criteria are also shown in the form of coefficients of correlation in Table I, columns e and j. There appears to be little difference between the two criteria in the magnitude of the biserial correlation coefficients obtained for each of the five predictor variables. The largest difference between the pairs of coefficients for any variable is 0.06 on the sociometric predictor, and this difference is smaller than the standard error of each coefficient. The difference may, therefore, be due to chance.

Since the pattern of correlation coefficients is so nearly the same for the two criteria, it is possible to consider the criteria as being perhaps different aspects of a single or general underlying criterion which may be described as "amenability to school goals," rather than an indicator of the two different categories of events postulated at the outset of this paper. Accordingly each predictor was examined against the combined criteria represented by three segments of the distribution, namely, the thirty-nine honor students, two hundred and sixteen children who were not on the honor roll but did complete high school, and the sixty-one children who left school prior to the completion of the twelfth grade. The triserial coefficients of correlation so obtained are shown in column k, Table I.

On the basis of the data reported in Table I, it appears that ratings made by teachers and ratings made by peers are better predictors of both honor roll attainment and of withdrawal from school than are the scores obtained on the standard commercial self-descriptive personality tests employed in the original study.

A review of the data, however, shows that important sex differences run through both the predictor and the criterion measures. When each of the triserial coefficients between the predictor and the general criterion is compared with the corresponding point-biserial coefficient for sex (obtained from the previous study and shown in column 1, Table I,) it appears that ratings have a larger correlation with sex than do the self-descriptions. It was also observed in the previous study (ibid., p. 40) that, "On the basis of correlational data, rating devices appear to have more in common with, or to be more valid measures of, the adjustment of boys as a group than of girls as a group; self-descriptive personality tests appear to have more in common with, or to be more valid measures of, the adjustment of girls as a group than of boys as a group."

Furthermore, both criteria employed in the present study are characterized by a pronounced sex differential. As shown in Table II, girls tended more often than boys to be recognized as high achievers: 18.8 per cent of the girls attained honor roll, whereas only 6.2 per cent of the boys did so. More girls (88.3 per cent) remained in school until graduation than did boys (73.4 per cent). The differences between both pairs of percentages are stable, as shown in Table II by a t-ratio of 3.4 for each. The phi-coefficient between sex and honor roll is 0.19 (with a potential maximum of 0.38); the phi-coefficient between sex and failure to graduate is also 0.19 (with a potential maximum of 0.49).

Accordingly, the data of Table I were recomputed for each sex separately and the results are shown in Tables III and IV. The correlations shown in Table IV show that the predictors function

TABLE II.—SEX DIFFERENCES IN ACHIEVEMENT

	Number					Per Cent			
		Honor roll*	Not on honor roll	Graduate	Non-graduate	Honor roll	Not on honor roll	Graduate	Non-graduate
Boys	162	10	152	119	43	6.2	93.8	73.4	26.6
Girls	154	29	125	136	18	18.8	81.2	88.3	11.7
Total	316	39	277	255	61				
Sex differences						12.6		14.9	
						t = 3.4		t = 3.4	

* Honor Roll of February, 1953

in different ways for the two sexes. Ratings for the sexes taken separately remain generally superior to self-descriptions in predicting the general criterion "amenability to school goals," but the raters' superiority is greater for the boys, whereas for girls it is only slightly better than the self-descriptions.

Teachers' ratings predict the general criterion better than do ratings by peers, and they predict the small proportion of honor roll boys best. The peers predicted withdrawal of girls relatively inefficiently, but they predicted honor roll standing about as well as teachers did. To some extent the lesser efficiency in prediction

TABLE III.—MEAN SCORES OF HONOR ROLL, NON-HONOR ROLL GRADUATES, AND NON-GRADUATE CHILDREN

	Boys					Girls				
	Honor roll	Non-honor graduate	Non-Graduate	Total	$\sigma_{\text{dis}_{\text{tot}}}$	Honor roll	Non-honor graduate	Non-graduate	Total	$\sigma_{\text{dis}_{\text{tot}}}$
N	10	109	43	162		29	107	18	154	
Forced choice	26.00	19.75	15.63	19.04	5.95	24.69	21.80	18.44	21.95	4.96
Sociometric	32.30	51.77	63.33	53.64	19.58	36.03	46.23	52.83	45.08	18.13
California self	71.20	69.35	66.42	68.69	11.22	74.90	70.51	62.89	70.45	12.41
California social	73.10	69.36	67.30	69.04	11.32	76.90	74.42	67.89	74.12	10.06
SRA basic difficulty	9.90	12.36	13.79	12.59	11.70	12.72	16.75	22.89	16.71	12.31

TABLE IV.—CORRELATIONS BETWEEN CRITERIA AND PREDICTORS BY SEX

Predictors	Criteria					
	Boys			Girls		
	Honor roll*	Non-graduate*	Com-posite†	Honor roll*	Non-graduate	Com-posite†
Forced choice	59	46	50	38	42	40
Sociometric	55	40	44	39	25	31
California self	11	16	16	25	36	30
California social	18	12	16	17	36	27
SRA basic difficulty	12	08	09	23	30	28

* Biserial r.

† Triserial r.

of withdrawing children by peers when compared with teachers' ratings of these children may be due to the fact that in the sociometric device the focus was predominantly upon selecting the better adjusted children.

The ratings account for almost all of the predicted variance in the two sub-criteria for boys. For girls, the ratings account for a relatively smaller (but still the major) part of the predicted variance and self-descriptive scores also contribute appreciably. The self-descriptive scores contribute particularly to the prediction of girls' withdrawal.

If we employ a correlation coefficient of 0.25 between California Self score and the Forced Choice score for girls, as derived in the previous study (ibid.), these two predictors then jointly yield a multiple R of 0.49 with girls' withdrawal, in comparison with the coefficient of 0.42 from teachers' Forced Choice ratings alone; the corresponding R for boys of 0.465 exceeds the coefficient of 0.46 (obtained from teachers' Forced Choice ratings) by an insignificant amount. It is evident, therefore, that the self-descriptive tests employed in this study added mainly to the prediction of withdrawal in the case of girls only.

From these data it appears that the conclusion quoted above from the previous study on the differential susceptibility of the two sexes to the two general categories of measurement is the overriding finding. Within this general finding, there is reason to believe that self-description will yield useful predictions of school withdrawal of girls. While these self-descriptive scores do not equal the teachers' Forced Choice scores in predictive efficiency for girls, they do contribute a significant increment to this predictive efficiency and they do not do so for boys.

The results of these computations are consistent with the conclusion of the previous study with respect to the greater validity of personality tests for predicting girls' adjustment than for predicting boys' adjustment. Ratings are generally more efficient than self-descriptions for predicting both aspects of maladjustment as defined in this study, with one important exception: ratings by *peers* are relatively ineffective in predicting withdrawal from school by *girls*. Personality tests are relatively useful supplements to ratings by teachers for predicting the withdrawal of girls. This refers only to ratings by teachers when obtained by means of a Forced Choice Test. As shown in the previous study, categorical ratings by teachers do not appear to differentiate significantly among girls.

The hypothesis stated at the outset of this paper is sustained in the following respects: Ratings are better predictors of honor roll status than of withdrawal *for boys only*; self-descriptive scores are better predictors of withdrawal than of honor roll status *for girls only*.

II

Although the self-descriptive test scores had the lowest relationship to manifest school adjustment, it was hypothesized that study of the large number of items employed in such instruments might yield some insight into the nature and causes of children's failure in adjustment. Accordingly, an item analysis of the SRA Youth Inventory was undertaken for a population of fifty-one boys who in the ninth grade had been rated as "seriously maladjusted" by their teachers. Of these fifty-one boys, twenty-five had dropped out of school before graduation, twenty-five completed the twelfth grade, and one had died.

An initial sorting of the Basic Difficulty scale of the Youth Inventory selected those items on which *both* sub-groups of twenty-five boys differed from the normative male group, that is, jointly either surpassed or fell below the normative males. Table V shows the items on which percentages of boys in the group designated as maladjusted by teachers differed from percentages of boys in the

TABLE V.—ITEMS ON WHICH "MALADJUSTED" BOYS DIFFER FROM THE NORMATIVE BOYS

		Normal	Malad-justed*	Differ-ences	t
	A. Items to Which the Normative Boys Respond More Often than the "Maladjusted" Boys				
94	I must always be "on the go"	30	18	12	2.14
125	I'm ill-at-ease at social affairs	26	10	16	3.64
	B. Items to Which the "Maladjusted" Boys Respond More Often Than the Normative Boys				
29	My teachers don't understand me	11	26	15	2.38
32	My teachers don't like me	5	16	11	2.12
72	I have trouble keeping my temper	27	44	17	2.39
76	I can't help daydreaming	29	46	17	2.36
84	I can't see much future for myself	11	30	19	2.92
106	I can't help feeling bad when I can't get my own way	6	26	20	3.17
110	I'm always thinking up alibis	15	28	13	2.03
113	I can't do anything right	2	16	14	2.69
158	I feel that I'm a burden on my parents	10	22	12	2.03
198	My parents neglect me	2	12	10	2.17
269	I feel that I'm not living up to my religion	22	38	16	2.29

normative group by a t-ratio of 1.96 or more, significant at the 0.05 level. Thirteen of the one hundred one items met this standard.

On the basis of data presented in Table V, A, it appears that the place of two items (94) and (125) in the Basic Difficulty Scale is open to question. Apparently more boys in the normative adolescent group described themselves as "I must always be 'on the go' " and "ill-at-ease at social affairs" than did maladjusted boys. Possibly being "on the go" is an acceptable or even desirable part of the adolescent stereotype; and it may be acceptable for an adolescent boy to attend a social affair and feel ill-at-ease, whereas maladjusted boys may avoid awareness of this experience by not attending at all.

The remaining items (Table V, B) which set maladjusted boys most sharply apart from the normative group appear to deal with difficulties in establishing control of impulses (items 72, 76, 106) and with a sense of failure to attain various social or school standards.

A second independent sorting selected items on the basis of differences between the two sub-groups of maladjusted boys. Because of the small size of these two sub-groups, it was not expected that differences in percentages responding to the items when so classified would be very reliable. Those items on which difference between the two groups showed a t-difference of 1.50 or more, indicating that differences of the observed magnitude could be expected to occur by chance in thirteen of one hundred instances, are shown in Table VI. Of the one hundred one items, fourteen satisfied this standard. The differences observed, therefore, may have arisen by chance and the discussion below is only suggestive of hypotheses to be tested with larger populations. Nevertheless, in view of the fact that the patterns of response appear to have coherence and the fact that the data represent a follow-up involving a three-year span, the items with largest t-scores for the differences between

TABLE VI.—ITEMS SHOWING DIFFERENCES BETWEEN TWENTY-FIVE MALADJUSTED GRADUATING AND TWENTY-FIVE MALADJUSTED WITHDRAWING BOYS

	Item	Graduating	Male Test Norm	Withdrawing	t-Ratio (G-W)
		Per cent	*Per cent*	*Per cent*	
6	I wish I could be more calm when I recite in class	56	51	32	1.76
11	I hate school	8	10	28	1.90
13	I wish I could quit school now	4	7	24	2.13
28	My teachers play favorites	32	21	12	1.75
71	I'm easily excited	32	14	8	2.22
87	I lack the drive others have	4	15	20	1.80
103	I'm afraid of failure or humiliation	24	17	8	1.58
119	I wish I were more popular	40	36	16	1.97
132	I need to learn to be a good listener	44	24	24	1.53
161	I feel there's a barrier between me and my parents	0	7	16	—
162	I can't discuss personal things with my parents	8	17	28	1.90
168	My parents play favorites	4	8	20	1.80
188	I feel like leaving home	8	8	24	1.58
230	I think about sex a good deal of the time	24	17	8	1.58
296	I wonder about the after-life	28	21	8	1.90

percentages are reported in Table VI for such use as they may have in formulating hypotheses for future study.

From Table VI it appears that the maladjusted boys who remained in school report that they feel impulse-ridden (items 71, 230); that they are anxious about failing to measure up (items 103, 296); and they express a need to make a change in self (item 132). These boys feel oriented *toward* adults: they notably do *not* suffer from feelings of inability to discuss personal things with parents (item 162) and they are highly concerned about the supposed partiality of their teachers (item 28). The maladjusted boys who withdrew from school felt that their parents played favorites (item 168); that there was a barrier between them and their parents (item 162). They were *not* concerned about teachers playing favorites (item 28). They felt that they lacked the drive others had (item 87), did not feel disturbed when required to recite in class (item 6), and did not aspire to greater popularity (item 119). They did not express fear of failure or humiliation (item 103) and could be described at the time of taking the test in the ninth grade as a defeated group who openly sought to get away from both home (item 188) and school (items 11 and 13).

SUMMARY

This paper has indicated that ninth grade teachers' Forced Choice Test ratings of children's adjustment are an indicator of children's amenability to school goals and are generally superior for this purpose to self-descriptive personality tests. Sex differences appear which are significant for practical prediction (i.e. selection) purposes. The pattern of correlations between the two categories of predictor (ratings and tests) and the two sub-categories of the general criterion (honor roll and withdrawal) is specific for each sex. Self-descriptive personality test scores appear to be relatively useful supplements to Forced Choice Test ratings by teachers for predicting withdrawal from school by girls.

It has also been suggested in this paper that the maladjusted category included both (1) those children who had difficulty in control of impulses and who were concerned about attaining approval through school performance, and also (2) those children who had given up the hope of ever obtaining approval for school effort. It has not been shown in this or in the previous study that these two types of children could not be differentiated by teachers.

The data have shown that it may be possible to key items on self-descriptive personality tests to distinguish between varieties of maladjustment which have different functional significance for public education and public health programs. In particular it may be possible to discover, either through carefully formulated questions to teachers or through experimentally derived keys for self-descriptive tests, which children would be readily amenable to the simple provision of therapeutic help and which children may require more aggressive casework techniques.

REFERENCES

1) Charles A. Ullmann, *Identification of Maladjusted School Children,* Public Health Monograph No. 7, Public Health Service Publication No. 211, Federal Security Agency, Washington, D. C., 1952.

DIFFERENTIAL SCORES AND UNSTABLE PERSONALITY

ROBERT C. SCARF

Ball State Teachers College, Muncie, Indiana

The widely used psychological examination for college freshmen, published by the American Psychological Association, is the basis of this report. The examination has two parts, a quantitative part and a linguistic part. Each freshman therefore has two scores, a Q-score and an L-score. The difference between the two is called an "unevenness" score. Most freshmen make scores that are fairly close, but a few show a big difference, being average in one score but close to the top, or the bottom, in the other while a rare case may be near the top in one, and near the bottom in the other. The median differential score, or unevenness score, is 16 percentile points, for one thousand nine hundred freshmen at Ball State Teachers College in 1948 and 1949.

Two groups of freshmen were compared. In one group the scores were even, or nearly so, the difference was less than six percentile points. In the other the difference was great, at least forty percentile points. In making up these two groups, certain records were excluded—freshmen who were over the age of twenty-one; who had attended another college for more than one year; who remained in school for less than two terms; and whose total score on the ACE test was ranked below the thirtieth percentile. This thirtieth percentile is regarded as a safe boundary. Those above it could surely get at least passing grades.

The data for men and for women were treated separately, and the data for freshmen who ranked higher in Q-score were treated separately from the data for those who ranked higher in L-score. No significant differences appeared; therefore, all data are included in one comparison.

The uneven group was found to do worse on the basis of grades. They were on probation, or disqualified for poor academic work more often. They earned more "F's", and more "D's." These facts are shown in Table I.

The uneven group finally graduated almost as many as the even group; forty-nine per cent as compared with fifty-three per cent, an insignificant difference. However, the uneven group took more

TABLE I.—ACADEMIC RECORD OF TWO GROUPS OF COLLEGE STUDENTS

Differential Score	N	Per Cent on Probation	Per Cent Disqualified	Per Cent Earning "D"	Per Cent Earning, "F"
Below 6	113	10.6	0.9	37.4	11.4
Over 40	180	25.0	7.2	56.1	25.0

All critical ratios are greater than 3.

TABLE II.—PERSISTENCE IN COLLEGE OF TWO GROUPS OF STUDENTS

Differential Score	N	Per Cent Graduating	Average Terms in Attendance	Per Cent Spending Excessive Terms to Graduate
Below 6	113	53	9.5	6.2
Above 40	180	49	7.6	30.6

time to graduate. 30.6% spent more than the usual twelve terms, while of the even group, only 6.2% spent more than twelve terms before receiving a degree. The even students spent more time in school, 9.5 terms, compared with 7.6 terms for the uneven group. Table II shows the data.

Another comparison is possible. One comparison, shown in Table I, is based on the number of students, i.e., of one hundred and thirteen students in the even group, only one was disqualified for poor scholarship, while thirteen out of one hundred and eighty students in the uneven group were disqualified. This does not tell the story, as the time spent in school is also a factor. That is, a student who spends six terms in school before disqualification has a better record, in a way, than one who is disqualified after three terms. This comparison, with terms in attendance as the basis, was made and the results are shown in Table III.

TABLE III.—ACADEMIC RECORD OF TWO GROUPS OF COLLEGE STUDENTS, WITH TERMS OF ATTENDANCE AS THE BASIS

Differential Score	Terms in Attendance	Per Cent of Terms on Probation	Per Cent of Terms Disqualified	Per Cent of "D's"	Per Cent of "F's"
Below 6	1072	1.5	0.1	11.4	1.7
Above 40	1372	9.5	1.2	22.5	5.8

All critical ratios are greater than 3.

Burea Edn.

From this table, it is seen that the uneven students were even worse than they seemed in Table I. It is surprising that the number of graduates is nearly the same for the even and the uneven groups. However, the uneven students often spent more than the usual twelve terms in order to overcome their deficiencies, and finally earn the degree.

Another way in which the two kinds of students differed was in the variation of grades received. The uneven students who received "D" at least once, also received "A's"; and the number of "A's" was as great as for the even students. Of the even students, forty-four per cent received at least one "A," while forty-eight per cent of the uneven students received at least one "A." It was concluded that the poor academic records were not due to lack of ability.

The two groups were also compared with respect to their total raw scores on the ACE. The result of this comparison is shown in Table IV. The data are given for each year separately, as different editions of the ACE test were used each year; in 1948 the 1946 edition, in 1949 the 1949 edition. The differences in the mean scores were insignificant. The conclusion is that the uneven students were substantially equal to the even students in general mental ability.

The large number of "D" and "F" grades were due to some other factor, presumably to defective personality.

To study another aspect of personality, all deans, counselors, and dormitory directors were asked to name the students who were regarded as queer or troublesome, or who had been involved in disciplinary action, or who, after leaving college, were known to have had serious difficulty. The psychosomatic cases all came from the writer's own file. Psychotics were listed who had been referred to a psychiatrist, or who were known to have been sent to a mental hospital.

TABLE IV.—THE TOTAL ACE SCORES OF TWO GROUPS OF STUDENTS

Year of Entrance	Kind of Student	N	Mean	Standard Deviation	Standard Error of the Difference
1948	Uneven	99	115.8	11.1	2.3
	Even	54	117.5	14.6	
1949	Uneven	81	107.0	10.8	1.8
	Even	59	108.8	11.1	

TABLE V.—DIFFERENTIAL SCORES OF UNSTABLE PERSONALITIES

Category	N	Median Differential Score	Per Cent Above Median of All Freshmen
Defective personality	78	35	96
Stutterers	5	29	80
Disqualified (non-academic)	5	43	100
Psychosomatic cases	22	33	74
Excluded from practice teaching	12	31	83
Suicide	12	31	92
Murderer	1	34	—
Frank psychosis	16	9	18

The list included students with psychosomatic complaints, general lack of adaptability, stealing, homosexuality, poor discipline, suicides, and one murder. Records below the thirtieth percentile on the ACE test were disregarded, as it was desired to exclude those whose difficulties were complicated by mediocre intelligence.

The differential scores of these people are shown in Table V. In general, the differentials are very high. It should be recalled that the median differential score is 16 for freshmen. For each category, the per cent of cases who exceed this, and therefore, can be regarded as abnormal, is given. These percentages are also high.

The list is admittedly incomplete. It depends in part on memories, going back for fifteen or twenty years. However, no one was included whose record was not adequate, showing good evidence of personality defect. Those excluded from practice teaching were excessively timid, or extremely rigid and severe, or had some other defect. Academic deficiency was not the cause of exclusion.

The finding for frank psychosis is surprising. These students were committed to mental hospitals, or were known to have consulted a psychiatrist. The low differential scores lend no support to the usual statement that psychosis is very like neurosis, only in more severe form.

CONCLUSION

The conclusion is that the psychological examination, often used with college freshmen, is a fairly good prognostic device. It is possible to use the examination to say which freshmen will have personality defects, or neurotic and unstable personalities, but it is useless as a predictor of psychotics.

Some theory which accounts for these facts is called for. Probably the best one is that the stable, well-balanced freshman attacks any task with equal zest and does his best on all tests. The unstable freshman, on the other hand, will do his best on one, but will be upset on another, resulting in uneven scores. The unevenness is therefore a sign of his total personality.

SEX AS A VARIABLE IN INTELLIGENCE TEST PERFORMANCE

MORONI H. BROWN and G. ELIZABETH BRYAN

University of Utah

Sex differences in intelligence test performance have been demonstrated in the standardization (*10, 13, 15*) and in later studies (*1, 2, 6, 8, 9, 11, 12*) of the three most widely used individual intelligence scales—the Stanford-Binet, the Wechsler-Bellevue and the Wechsler Intelligence Scale for Children (WISC). The differences have occurred even though, in the construction of these scales, tests with minimal sex differences were selected. Most writers have been hard-put to explain these differences, especially when the intelligence scales are supposed to yield comparable results for males and females. As Seashore, Wesman, and Doppelt (*10*) have noted, there seem to be three major possibilities which logically could account for sex differences. These are: (1) that the tests in the scales are unbiased and one sex actually is superior in performance on the tests; (2) that the sexes are equal in mental ability, but the chosen test items turned out to be slightly biased in favor of one sex; or (3) again assuming that general ability is not sex differentiated, that the sampling of one of the sexes was somehow chosen with a slight bias.

Consequent to the rigorous avoidance of actually hypothesizing sex differences in intelligence test performances, the published studies often carry a small apologetic sentence tucked away in the conclusion section to the effect that the observed differences were possibly due to sampling bias or to test bias.

In the development of these individual intelligence scales, the authors have emphasized the absence of sex differences in the total scales but all three scales contain component tests which favor one or the other sex. For example, in the 1937 revision of the Stanford-Binet Scales, Terman and Merrill (*13*) eliminated several tests which yielded the largest sex differences, but they acknowledged that a considerable number of the tests which were included in the revised scales showed statistically significant differences in the percentages of success for males and females. These differences are supposed to largely cancel out when the tests are combined, but even so, on the Full Scale, the authors reported a two-point superiority for boys between the ages of six and thirteen years. Above the age of fourteen, there was a four-point superiority for boys. These sex differences were obtained on their standardization group of two thousand four hundred and forty-four subjects ranging from five to eighteen years of age. Their comment on the observed sex differences was that there was a possibility that some of these differences were due to biased sampling.

Terman and Tyler (*12*, p. 1068) report a study of Scottish school children with the 1937 Stanford-Binet Scale in which sex differences of 3.24 IQ points were obtained in favor of the males. Their comment on these results was that the direction and magnitude of sex differences in intelligence is primarily dependent upon the test used. In other words, it was due to a test bias.

Seashore et al. (*10*) in an article on the standardization of the Wechsler Intelligence Scale for Children, found that girls excelled boys at the ages of five, six, seven, and nine on the Verbal, Performance, and Full Scale IQ scores, but that at the age of eight and from the ages of ten to fifteen, boys excelled girls on all three IQ scores. The greatest differences were in the Verbal scores at the age of eight and from ten to fifteen where the boys' superiority ranged from three to five IQ points. Although the significance of these sex differences was not given in the article, they were computed by the present writers and some of them were found to be statistically significant (0.05 level).

Wechsler (*15*) stated that there were no statistically significant sex differences in total scores for the group on which the Wechsler-Bellevue Scale was standardized, but that women tended to have higher mean total scores at almost every year level. This statement seems to imply that there might have been statistically significant differences on some or all of the tests in the Scale. There is some question about what Wechsler was referring to when he said that "these differences in themselves do not prove anything, but taken together with other evidence, have led this writer (that is, Wechsler) to the belief that it may be possible to demonstrate a measurable superiority of women over men so far as general intelligence is concerned." He then went on to say that he did not have the necessary data as yet to back up this generalization (even though he had just mentioned that "other evidence" had contributed to his belief).

Despite Wechsler's attempt to construct a scale in which no total score differences would obtain, such differences have been reported in later studies. Strange and Palmer (*11*) noted statistically significant sex differences favoring males on nine of the subtests and on all three IQ scores. In their study, these authors used the test records of a group of one hundred and forty-five male and ninety female adult psychiatric out-patients whose mean age was in the early thirties. Contrary to Wechsler's statement about the superiority of women in intelligence, the Strange and Palmer study showed higher scores for males which were significant at the 0.001 level for Verbal and Full Scale IQ's and at the 0.02 level for Performance IQ's. These writers mentioned that theirs was an exploratory study, but even so the sex differences which were observed in a sample as large as theirs should serve as a warning that sex differences should be checked in further studies which use the Wechsler-Bellevue Scale since such differences might change the interpretation of the results.

Following the Strange and Palmer report on sex differences, Goolishian and Foster (*6*) used a collection of test records of three hundred and ninety-two adult psychiatric male and female patients whose mean age was about thirty years. The results showed sex differences significant at the 0.05 level or better favoring males on five of the subtests and on all three IQ's. The sex differences in Verbal IQ's were significant at the 0.001 level, the Performance IQ differences were significant at the 0.05 level, and the Full Scale IQ differences were significant at the 0.01 level.

A later study (*2*) used a group of four hundred non-hospitalized and non-referred subjects from a public school and college population. There were forty subjects of each sex at five age levels. The results showed sex differences at all five age levels favoring males on the Full Scale weighted scores but none of them was significant. However, on the Verbal Scale, males had higher weighted scores than did females and at three age levels the differences were significant at or beyond the 0.05 level. There were less pronounced differences on the Performance Scale but at four of the five age levels, female subjects had weighted scores equal to or higher than males. The two oldest age ranges showed sex differences significant at the 0.025 and 0.05 levels, respectively.

Although the above mentioned studies using the Wechsler-Bellevue all indicated sex differences generally favoring males, it should be noted that the samples of subjects in these studies were not similar. The first two studies used the test records of psychiatric subjects exclusively while the third study included only non-clinic and presumably "normal" subjects. In addition, the first two studies used unequal numbers of male and female subjects. More importantly, only the mean ages were reported for the psychiatric groups and no information regarding the age ranges or age distributions was given. This is a serious limitation since intelligence test performance has been shown to vary with age (*1*) so that one cannot assume constancy at different age levels.

To the extent that the Stanford-Binet, the Wechsler-Bellevue, and the WISC are valid measures of intelligence, the studies here quoted all indicate that there are sex differences in the levels of over-all intellectual functioning. However, it seems that a pertinent question in regard to sex differences concerns the rationale for attempting to eliminate differences in performance between the sexes on intelligence tests when such differences are readily accepted in areas which deal with physiological and emotional reactions, vocational aptitudes, interest and personality inventories, etc. Somehow the possibility of sex differences in intelligence test performance seems to have been an unacceptable hypothesis. This resistance apparently is related in part to a conviction that sex differences could not exist because this would imply that one sex is superior to the other. Consequently, the explanations of any demonstrable sex differences usually have incorporated either test or sampling bias as the major contributing factor. In contradiction to this approach, it would seem better to stress that qualitative differences may exist and, particularly that they do, in fact, exist in the currently used intelligence scales so that intelligence as it is measured is not equal for the sexes regardless of whatever inherent differences may or may not be present. If the "amount" of intelligence is not the issue but rather the "kind of intelligence in test performance," then it should be possible to retain a test in a scale as a useful measuring tool as long as male and female norms are established for that scale. It might be noted, incidentally, that for the Cornell-Coxe Performance Ability Scale (*5*) there is one table of smoothed IQ scores for the combined male and female standardization subjects at different age levels, but separate tables of norms for male and female subjects are presented and they show a definite tendency for males to obtain higher scores at all age levels.

Perhaps the individual scales which are being used at the present time would be more useful if, in their standardization, the component tests which yielded different results for males and females had been retained. Inasmuch as the differences between the performance of male and female subjects vary with age, it would seem that the standardization norms would be more valuable if they were reported for discrete sex-by-age groups. The individual treatment of age and sex is not sufficient since both variables interact and their relationships are not constant. This inconsistency of the age and sex variables has been indirectly demonstrated by the lack of agreement among the results of various studies which have used samples of different age ranges (*3*, *4*, *7*, *8*, *9*, *12*, and *14*). Direct evidence for age-by-sex interaction on the Wechsler-Bellevue has been reported (*1*) and it was shown that not only did the Verbal, Performance, and Full Scale scores vary for each sex at different ages but also that the weighted scores of the component subtests varied in their rank order from one age level to the next.

By recognizing and allowing for sex differences, it might be possible to have a psychometric device which would prove to be more sensitive to unusual or pathological intellectual functioning. As it is, however, no such general intelligence scale exists. Since sex differences are present in the currently used scales, then the recognition of these differences at least should throw some doubt onto the conservative and glib generalization that intellectual abilities are equally distributed between the sexes.

REFERENCES

1) M. H. Brown, and G. Elizabeth Bryan, "Age and Sex Variability on the Wechsler-Bellevue Scale," Paper read at Western Psychological Association, Long Beach, May, 1954.

2) M. H. Brown, and G. Elizabeth Bryan, "Sex Differences in Intelligence." *Journal of Clinical Psychology*, 11: 303–304, 1955.

3) G. Elizabeth Bryan, *A Study of the Performance of Adolescents with Diagnosed Organic Brain Damage on the Wechsler-Bellevue Intelligence Scale, Form I*, Unpublished Master's thesis, University of Utah, 1954.

4) G. Elizabeth Bryan, and M. H. Brown, "Levels of Difficulty for Adolescents on the Wechsler-Bellevue Scale," Paper read at Western Psychological Association, Long Beach, May, 1954.

5) Ethel L. Cornell, and W. W. Coxe, *A Performance Ability Scale: Examination Manual*, New York, World Book Co., 1934.

6) H. A. Goolishian, and A. Foster, "A Note on Sex Differences on the

Wechsler-Bellevue Test." *Journal of Clinical Psychology*, 10: 298–299, 1954.

7) W. H. Guertin, "The Effect of Instructions and Item Order on the W-B Arithmetic Subtest," *Journal of Genetic Psychology*, 85: 79–83, 1954.

8) R. J. Howell, "Sex Differences and Educational Influences on a Mental Deterioration Scale," *Journal of Gerontology*, 10: 190–193, 1955.

9) R. D. Norman, "Sex Differences and Other Aspects of Young Superior Adult Performances on the Wechsler-Bellevue," *Journal of Consulting Psychology*, 17: 411–418, 1953.

10) H. Seashore, A. Wesman, and J. Doppelt, "The Standardization of the Wechsler Intelligence Scale for Children." *Journal of Consulting Psychology*, 14: 99–110, 1950.

11) F. B. Strange, and J. O. Palmer, "A Note on Sex Differences on the Wechsler-Bellevue Tests," *Journal of Clinical Psychology*, 9: 85–87, 1953.

12) L. M. Terman, and Leona E. Tyler, "Sex Differences in Abilities," In: L. Carmichael, *Manual of Child Psychology*, pp. 1064–1114, New York, John Wiley and Sons, 1954.

13) L. M. Terman, and Maude A. Merrill, *Measuring Intelligence*, New York, Houghton Mifflin, 1937.

14) Julia R. Vane, and Virginia W. Eisen, "W-B Performance of Delinquent and Non-Delinquent Girls," *Journal of Consulting Psychology*, 18: 221–225, 1954.

15) D. Wechsler, *The Measurement of Adult Intelligence*. (Third Edition). Baltimore, The Williams & Wilkins Co., 1944.

THE INTERRELATIONSHIP OF CERTAIN LINGUISTIC SKILLS AND THEIR RELATIONSHIP WITH SCHOLASTIC ACHIEVEMENT WHEN INTELLIGENCE IS RULED CONSTANT

EDWIN E. VINEYARD

Director of Counseling

and

HAROLD W. MASSEY

Education Department Chairman

Panhandle Agricultural and Mechanical College, Goodwell, Oklahoma

PURPOSE OF THE STUDY

Previous research has frequently revealed that various linguistic skills such as vocabulary, reading comprehension, and spelling are positively related and that each is related to scholastic achievement. Previous research has also revealed that each of these linguistic skills is positively related to intelligence of the symbolic type as measured by the usual group intelligence test. However, little research has been forthcoming concerning the degree of relationship remaining among these linguistic skills and between them and scholastic achievement when the intelligence factor with which each appears to have some degree of saturation is ruled constant experimentally or statistically. This is the problem investigated in the present research.

Specifically, the present study seeks to identify the following relationships:

(1) The coefficients of correlation between objective measures of vocabulary, speed of paragraph comprehension, and spelling.

(2) The coefficients of correlation between intelligence scores and objective measures of vocabulary, speed of paragraph comprehension, and spelling.

(3) The coefficients of correlation between college grade point

averages and objective measures of vocabulary, speed of paragraph comprehension, spelling, and intelligence.

(4) Partial coefficients of correlation between each of the following with intelligence ruled constant: vocabulary and speed of paragraph comprehension, spelling and speed of paragraph comprehension, and spelling and vocabulary.

(5) Partial coefficients of correlation between each of the following and college grade point averages with intelligence ruled constant: vocabulary, speed of paragraph comprehension, and spelling.

MEASUREMENT OF VARIABLES

The Nelson-Denny Reading Test, Form A, yields objective measures of both vocabulary and speed of paragraph comprehension. These scores were used as measures of these abilities in the present research. The spelling test of the Differential Aptitude Tests was used as a measure of ability in spelling. Total scores on the 1952 edition of the American Council on Education Psychological Examination were used as measures of intelligence. Each of these tests was administered as a part of the freshman testing program at Panhandle Agricultural and Mechanical College. Grade point averages used as criteria in this investigation were first semester averages. The subjects for the present investigation numbered one hundred seventy six and represented the entire 1955–56 freshman class of the above college who remained for one semester and for whom complete test results were available.

RESULTS OF THE STUDY

In attacking the problem, the first step was that of ascertaining the relationships present among the three linguistic skills with which the study was concerned. These coefficients of correlation are shown in Table I.

The coefficient of correlation between vocabulary and speed of

TABLE I.—RELATIONSHIP BETWEEN PAIRS OF LINGUISTIC SKILLS

Pairs of Skills	r
Vocabulary and speed of paragraph comprehension	0.621
Spelling and speed of paragraph comprehension	0.408
Spelling and vocabulary	0.668

paragraph comprehension was 0.621, indicating a substantial positive relationship. The correlation between vocabulary and spelling was 0.668, also indicating a substantial positive relationship. Between spelling and speed of paragraph comprehension the correlation was 0.408, indicating a moderately strong relationship. Thus, the present results agree with those usually found in similar studies in that these three linguistic skills exhibit definite positive interrelationships. Under such conditions one may legitimately hypothesize the possibility that at least a part of this interrelationship may be due to the saturation of each with some common factor, and proceed to test such a hypothesis. The present study hypothesizes that intelligence is such a common factor present in each of these separate abilities and seeks to ascertain what relationships between these skills remain when the influence of this common factor is removed. Table II shows the relationships found between each of the three skills and intelligence.

From the data in Table II it is seen that each of the three linguistic skills measured is strongly related to intelligence, the coefficients of correlation being 0.708, 0.659, and 0.573 between intelligence and vocabulary, speed of paragraph comprehension, and spelling, respectively. These data seem to support the premise that intelligence may be a common factor accounting at least partially for the strong relationship existing among the three variables. The next logical step in the testing of such a hypothesis is the partialling out of that portion of relationship between pairs of linguistic skills due to the common factor of intelligence and identifying the degree of relationship remaining. The results of this process are shown in Table III in the form of partial r's.

The correlation coefficients shown in the above table are those between pairs of linguistic skills with intelligence ruled constant statistically. By referring to both Table I and Table III one may observe the decrements in degree of relationship between these

TABLE II.—RELATIONSHIPS BETWEEN LINGUISTIC SKILLS AND INTELLIGENCE

Linguistic Skill	r
Vocabulary	0.708
Speed of paragraph comprehension	0.659
Spelling	0.573

TABLE III.—RELATIONSHIPS BETWEEN PAIRS OF LINGUISTIC SKILLS WITH INTELLIGENCE RULED CONSTANT

Pairs of Skills	Partial r
Vocabulary and speed of paragraph comprehension	0.290
Spelling and speed of paragraph comprehension	0.049
Spelling and vocabulary	0.452

pairs of skills when the influence of intelligence is removed. Thus, the correlation between vocabulary and paragraph comprehension drops from an original 0.621 to a modest 0.290. The correlation between spelling and speed of paragraph comprehension drops from an original 0.408 to a statistically insignificant value of 0.049. The correlation between spelling and vocabulary dropped from a strong 0.668 to a moderate value of 0.452.

It is seen, then, that the removal of the common factor of intelligence very definitely reduces the relationships among these linguistic skills. The relationship affected most by this removal appeared to be that between spelling and speed of paragraph comprehension, this relationship apparently being almost entirely accreditable to the relationship of both factors with the third variable of intelligence. Also considerably affected was the relationship between vocabulary and speed of comprehension. It would appear that much of the strength of the relationship between these two abilities is attributable to intelligence since the correlation coefficient is considerably reduced by its removal. However, all of the relationship between these two variables is not due to a saturation of both with the intelligence factor since there remains a definite positive relationship when its influence is removed. The observed decrement in degree of relationship is not so great between spelling and vocabulary as those observed among other pairs of linguistic skills when intelligence is ruled constant statistically. Although there is a reduction in the coefficient of correlation when intelligence is partialled out there remains a fairly substantial relationship.

The study was further concerned with the relationship of all variables with scholastic achievement and with the relationship remaining between each of the three linguistic skills with scholastic achievement when the variable of intelligence is ruled constant. Table IV shows the relationship found between each of the

TABLE IV.—RELATIONSHIP OF ALL VARIABLES WITH COLLEGE GRADE AVERAGE

Variable	r
Vocabulary	0.513
Speed of paragraph comprehension	0.433
Spelling	0.557
Intelligence	0.518

studied variables and college success in terms of grade point average.

From the data in Table IV it may be seen that all of the variables studied have strong coefficients of correlation with scholastic success. The relationship between intelligence as measured by the ACE and scholarship is similar to that usually found by other researchers. Those coefficients of vocabulary and spelling with grade average are surprisingly high considering both the brevity and the limited scope of the tests of these abilities in comparison with the longer and more comprehensive ACE.

When that part of the relationship between the three linguistic skills and scholarship which is attributable to the common variable of intelligence was partialled out, the data in Table V were obtained. The degree of remaining relationship is shown in the form of partial r's.

The partial r of 0.144 remaining between speed of paragraph comprehension and scholastic average with intelligence ruled constant is not statistically significant. The partial r's between vocabulary and spelling and grade average are significant at the 0.01 level of confidence. In all cases there is a decrement in relationship observed between the linguistic skills and scholarship when the influence of intelligence is removed. However, there remains a definite positive relationship for both spelling and vocabulary with scholastic success.

TABLE V.—RELATIONSHIP BETWEEN LINGUISTIC SKILLS AND GRADE AVERAGE WITH INTELLIGENCE RULED CONSTANT

Linguistic Skill	Partial r
Vocabulary	0.245
Speed of paragraph comprehension	0.144
Spelling	0.373

SUMMARY AND CONCLUSIONS

From the results of the present study one may safely make the following generalizations:

(1) The linguistic skills of vocabulary, paragraph comprehension, and spelling are related substantially with one another, and all are related substantially with intelligence. Each of these variables is also related substantially with scholastic success.

(2) Much of the strength of the relationships present among these linguistic skills is due to a common saturation with intelligence. Much of the strength of the relationships present between each of these linguistic skills and scholarship is due to their common saturation with intelligence.

(3) Although the removal of the influence of intelligence reduces the relationship between vocabulary and paragraph comprehension, there remains a relationship sufficiently strong to justify attempts to improve reading comprehension through vocabulary training.

(4) There is no significant relationship between speed of paragraph comprehension and spelling ability when the influence of intelligence is removed. This would seem to indicate that the slow reader is not necessarily a better speller than the fast reader of equal intelligence.

(5) Although the removal of the influence of intelligence somewhat reduces the relationship between vocabulary and spelling ability, there remains a fairly strong relationship between these skills.

(6) There is no significant relationship between speed of paragraph comprehension and scholarship when the influence of intelligence is removed. This would seem to indicate that programs in speeded reading can do little to influence the quality of scholarship although they may serve to reduce considerably the time spent in acquiring a given scholastic proficiency through reading.

(7) Although the removal of the influence of intelligence reduces the relationship of vocabulary and spelling with scholastic achievement, there remains a significant relationship between these skills and scholarship. This would seem to indicate that reading programs which concentrate on vocabulary development and programs in remedial spelling offer definite possibilities for improvement of general scholarship.

(8) Research which goes no further than the computation of

zero order relationships between symbolic skills is incomplete and may actually be misleading. Future research in this area must recognize the possibility of common saturations and explore these possibilities in order to be scientifically complete.

REFERENCES

Guy L. Bond, and Theodore W. Clymer, "Interrelationship of the SRA Primary Mental Abilities, Other Mental Characteristics, and Reading Ability," *Journal of Educational Research,* Vol. 49, No. 2, pp. 131–136, October, 1955.

Harry Grater, and W. A. Thalman, "A Statistical Analysis of the Relationship Between American Council on Education Psychological Examination Ratings and Grade-Point Averages," *Journal of Educational Research,* Vol. 49, No. 4, pp. 307–310, December, 1955.

W. S. Gray, *Summary of Investigations Relating to Reading,* Supplementary Educational Monographs, No. 28, University of Chicago Press, 1925. A summary of the results of 436 scientific studies published prior to July 1, 1924.

W. S. Gray, "Annual Summary of Investigations Relating to Reading:"
a. July 1, 1924 to June 30, 1925, *Elementary School Journal,* 26: 449–; 1926.
b. July 1, 1925 to June 30, 1926, *Elementary School Journal,* 27: 456–; 1927.
c. July 1, 1926 to June 30, 1927, *Elementary School Journal,* 28: 443–; 1928.
d. July 1, 1927 to June 30, 1928, *Elementary School Journal,* 29: 443–; 1929.
e. July 1, 1928 to June 30, 1929, *Elementary School Journal,* 30: 450–; 1930.
f. July 1, 1929 to June 30, 1930, *Elementary School Journal,* 31: 531–; 1931.
g. July 1, 1930 to June 30, 1931, *Elementary School Journal,* 32: 447–; 1932.
h. July 1, 1931 to June 30, 1932, *Journal of Educational Research,* 26; 401–; 1933.
i. July 1, 1932 to June 30, 1933, *Journal of Educational Research* 27: 564–; 1934.
j. July 1, 1933 to June 30, 1934, *Journal of Educational Research,* 28: 401–; 1935.
k. July 1, 1934 to June 30, 1935, *Journal of Educational Research,* 29: 407–; 1936.
l. July 1, 1935 to June 30, 1936, *Journal of Educational Research,* 30: 553–; 1937.
m. July 1, 1936 to June 30, 1937, *Journal of Educational Research,* 31: 401–; 1938.
n. July 1, 1937 to June 30, 1938, *Journal of Educational Research,* 32: 481–; 1939
o. July 1, 1938 to June 30, 1939, *Journal of Educational Research,* 33: 481–; 1940.
p. July 1, 1939 to June 30, 1940, *Journal of Educational Research,* 34: 401–; 1941.
q. July 1, 1940 to June 30, 1941, *Journal of Educational Research,* 35: 401–; 1942.
r. July 1, 1941 to June 30, 1942, *Journal of Educational Research,* 36: 401–; 1943.

s. July 1, 1942 to June 30, 1943, *Journal of Educational Research*, 37: 401–; 1944.

t. July 1, 1943 to June 30, 1944, *Journal of Educational Research*, 38: 401–; 1945.

u. July 1, 1944 to June 30, 1945, *Journal of Educational Research*, 39: 401–; 1946.

v. July 1, 1945 to June 30, 1946, *Journal of Educational Research*, 40: 401–; 1947.

w. July 1, 1946 to June 30, 1947, *Journal of Educational Research*, 41: 401–; 1948.

x. July 1, 1947 to June 30, 1948, *Journal of Educational Research*, 42: 401–; 1949.

y. July 1, 1948 to June 30, 1949, *Journal of Educational Research*, 43: 401–; 1950.

z. July 1, 1949 to June 30, 1950, *Journal of Educational Research*, 44: 401–; 1951.

aa. July 1, 1950 to June 30, 1951, *Journal of Educational Research*, 45: 401–; 1952.

bb. July 1, 1951 to June 30, 1952, *Journal of Educational Research*, 46: 401–; 1953.

cc. July 1, 1952 to June 30, 1953, *Journal of Educational Research*, 47: 401–; 1954.

dd. July 1, 1953 to June 30, 1954, *Journal of Educational Research*, 48: 401–; 1955.

ee. July 1, 1954 to June 30, 1955, *Journal of Educational Research*, 49: 401–; 1956.

E. B. Huey, *The Psychology and Pedagogy of Reading*, Macmillan, 1912, pp. 1–184. A summary of the results of fifty years of reading research in the United States and abroad.

Albert J. Kingston, and Clay E. George, "The Effectiveness of Reading Training at the College Level," *Journal of Educational Research*, Vol. 48, No. 6, pp. 467–471, February, 1955.

Southwest Reading Conference, *Explaining the Goals of College Reading Programs*, Yearbook V, Texas Christian University, Fort Worth.

Southwest Reading Conference, *Evaluating College Reading Programs*, Yearbook IV, Texas Christian University, Fort Worth.

Southwest Reading Conference, *What Colleges and Universities Are Doing in Planning and Improving Reading Programs*, Yearbooks I, II, & III, Texas Christian University, Fort Worth.

A. E. Traxler, and Agatha Townsend, *Ten Years of Research in Reading*, No. 32, Educational Records Bureau, New York, 1941. Reports on more than 600 studies covering the period 1930–39.

A. E. Traxler, *Another Five Years of Research in Reading*, Educational Records Bulletin, No. 46, Educational Records Bureau, New York, 1946. Reviews more than 500 studies for the period 1940–45.

A. E. Traxler, *Eight More Years of Research in Reading*, Educational Records Bulletin, No. 64, Educational Records Bureau, New York, 1955. Summarizes 760 studies published from 1945 through 1952.

AN EVALUATION OF TWO TYPES OF KINDERGARTEN ATTENDANCE PROGRAMS[1]

EVA DREIKURS FERGUSON

Northwestern University

In line with the growing trend in preschool education to decrease the number of hours of kindergarten attendance per week, the center in which the present study was made established a program of spaced attendance modified from a suggestion made by Gesell and his co-workers (*2*). The Yale group suggested that the assimilation of the newly learned skills gained through a kindergarten experience was provided to a greater extent by a spaced than by a continuous attendance program, especially for the three-year-old. Additional advantages allowed by spaced attendance were that it provided opportunity for mother and child to have outings together on the days the child was home, and that it would be less fatiguing for the child to have the extra days of nonattendance. The modified spaced attendance initiated at the center in which this study was made consisted of five mornings of attendance every two weeks, the children attending three days one week and two days the next week. Although this program ran successfully for more than two years, other preschool centers in Australia were reluctant to follow in this venture until it could be adequately demonstrated that spaced attendance was not depriving the child of the social and psychological development that was currently provided by daily attendance.

The purpose of the present study was therefore an evaluation of the spaced attendance program, by testing whether any noticeable differences could be found between daily and spaced attendance, with regard to the children's social, psychological and physical development. Specifically, the two attendance programs were compared on the following points: (a) the individual child's behavior, as displayed in free-play periods; (b) social cohesion of the total peer group; (c) physical development of the individual child; (d) attendance records and reasons for absences; (e) the individual child's behavior and development at home, as reported by the mother. The study covered a six-month period, from March to August, which included the fall and winter seasons in Australia.

---

[1] The present study was made in partial fulfillment for the Master's degree in the Psychology Department, Melbourne University. The data were gathered at the Lady Gowrie Child Center in Melbourne, Australia. Because no uniform terminology exists in the literature, the terms "preschool center" and "kindergarten" are used interchangeably in the present study.

METHOD

Groups

The children used in the study were between the ages of three and five. They came from a homogeneous, working-class neighborhood, with over seventy per cent of the parents having eight years or less of education. (Since in Melbourne the age limit for compulsory education is fourteen years, this sample is typical of the Melbourne working class.) Fifty-six children were selected from the available kindergarten population of the center. The majority of them (forty out of the fifty-six) had attended the center the previous year. Two groups, each of twenty-eight children, were formed by matching Ss in pairs on the basis of the following variables.

(1) Age, as measured in half-year units.

(2) Type of family background. This included: (a) the type of discipline and its degree of consistency used on the child in the home; (b) the degree of manageability of the child at home; (c) amount of family harmony; (d) sibling status, measured according to whether the child was an only child, had siblings close in age, or had siblings six or more years older than himself.

(3) Adjustment to peers and to kindergarten routine at the end of the previous kindergarten year.

The data on the family background was obtained from detailed records which are routinely kept for all the children at the center by a social worker. Those children who had attended the center the previous year were rated on a 3-point scale on the variables delineated under home background. The reliability of the various ratings was checked against the director's judgment regarding each child's home background, her judgment being based on her personal contacts and home visits with the family. This reliability was found to be high, agreement being above eighty per cent.

Adjustment to peers and to kindergarten was based on a rating scale used by the teachers. At the end of the previous year, the teachers rated all the children at the center who were to continue at the kindergarten for the year of the study.[2]

The children who had not attended the center the previous year were matched only on age and sex, as no adequate information on their home background was available. As far as possible, Ss were assigned randomly to the two groups, although completely random assignment was not possible. (In cases of siblings it was not possible, as it was desirable to have siblings in different groups, and in the few cases where matching of the pair was not complete, the children were so assigned as to keep the distribution of the matching variables equal in the two groups.) After the groups were formed, the director of the kindergarten, without knowledge of which children were in which group, arbitrarily designated one group as the spaced attendance and the other group as the daily attendance group. Thus, any systematic bias, of which child attended which type of program, was eliminated. The two groups will henceforth be designated as the S-A group, for spaced attendance, and the D-A group, for daily attendance. The S-A group attended Tuesdays and Fridays and alternating Wednesdays; the D-A group attended every day. The hours of attendance for both groups were nine to twelve for the first three months and nine-thirty to twelve-thirty for the last three months, the latter hours being part of the center policy to allow the children to sleep longer on the colder and darker winter mornings.

Two teachers were used in each group, as was the custom at the center. These were matched for length and type of teaching experience. The rooms Ss used were comparable in facilities, and Ss shared a common playground and playground equipment. Teaching programs and methods of handling the children were synchronized in conference with the four teachers, so that the children had comparable teaching experiences.

Because of unavoidable loss of Ss during the course of the study, the final number of matched pairs was twenty instead of the original twenty-eight. Some Ss remained in spite of the matched partner dropping out, however, and these Ss were included in some of the data. Thus, although there were only twenty pairs, the total number of Ss in the two groups was forty-eight. The losses were spread fairly evenly throughout the age and sex categories, so that the final age and sex distribution of the twenty matched pairs was almost comparable to the original group of twenty-eight. Of the final twenty pairs, nine were boys' pairs and eleven were girls', the mean age for both sexes, in March, being four years and one month. This is in contrast to the original group of fourteen boys' and fourteen girls' pairs whose mean age was three years and eleven months.

---

[2] Complete information regarding the matching procedure of the two groups can be found in the original Master's thesis. The matching was undertaken merely to safeguard comparability of the two groups at the onset of the study. The data indicate the matching to have been effective (see Table I), since a significant difference between the groups, in March, was found for only one item in the behavioral observations.

Types of Measurements Used

Behavioral observations. To assess Ss' behavior, an observation sheet with seventeen categories was devised. Figure 1 presents some of the major categories and a few illustrative items.[3] Each S was observed by the time-sampling technique for three minutes twice a week, during free-play periods. The S-A group was observed on Tuesdays and Fridays, the D-A group on Mondays and Thursdays. Ss were observed in random order, the matched partners being observed in the same relative order within their respective group. Different random lists were used for the six months of the study, so that only by chance would a given S be observed at the same time on several observation days. The three minute observation period was divided into six sections of thirty seconds each. The behavior shown by S at the time of observation was fully recorded by tallies in the appropriate categories.[4] In some cate-

---

[3] The complete Behavior Observation Categories Instrument may be obtained on written request to the author. For a detailed account of the categories and of their operational use the reader is referred to the original Master's thesis.

[4] Although there was no reliability check on the observations, any bias in observation that may have occurred would be suffered equally by the two groups. This is so, since the data were not analyzed until the end of the study, so that during the course of the observations the investigator did not know which way the results were going. The reason no reliability check was made was that no second observer was available for the vigorous training required for the use of the observation sheet. Admittedly, the lack of a reliability check on the observations is a methodological weakness of this study.

FIGURE 1

(1) Type of activity
    (a) Aimless or idle preoccupation
    (b) Looking on at some activity
    (c) Solitary activity
    (d) Parallel activity
    (e) Associative play activity
    (f) Coöperative play activity
    (g) Contact with adult
(2) Amount of serious absorption (constructiveness) during play
    (a) Passive sitting or onlooking
    (c) Mostly conversing rather than playing
    (f) Serious absorption and oblivion to surroundings
(3) Verbal communication during play
    (a) Primitive verbal usage
    (d) Talking to self or making noises
    (k) Arguing
(4) Degree of follower-leadership shown toward peers
    (b) Seeking or asking for help
    (e) Leading and/or organizing others
(7) Aggression towards peers: 1. no immed. contact; 2. act of offense; 3. act of defense.
    (f) Disrupting group or solitary play
(9) Crying
    (b) When hit or otherwise attacked by another
(1) Themes—Dramatic Play (With Imaginative Content and Rôle-Playing)
    (c) Very elaborate home situations
(4) Materials used
    (c) Complex (improvising to achieve greater realism)

FIGURE 1. Some of the major categories and illustrative items from the behavior observation categories instrument.

gories a tally meant that the checked behavior was shown predominantly in the thirty seconds, as for instance in category 1, "Type of activity," while in other categories a check was made for the behavior regardless of its duration in the thirty seconds, as for example in category 9, "Crying." Though only one S was focused upon at a time, the names of Ss with whom he interacted were noted on the side.

The data used for analyzing the results were derived as follows. For each item, a S's mean score for the month was obtained by dividing the total score for the month by the number of days the S was observed in that month. The possible range of scores for a given observation day was 0 to 6, a score representing the number of thirty-second intervals in which an item was checked. Thus, a score of 5 meant the behavior was shown in five of the six thirty-second intervals. If S was present for only two observation days in the month, and on one day his score was 0 and on the other day 5, his mean score for that item for that month would be 2.5. Observations were made about eight times a month, each group being observed twice a week. Since some Ss were occasionally absent, the number of observation days for a given S for one month ranged from two to eight days.

Sociometric data. To assess the social cohesion of the two groups, two sociometric tests were administered, one in late April and one at the end of July. The questions for each test were asked by a volunteer worker, who asked each S whom he likes to play with in three specified situations: in the sandpit; at clay; and with blocks (for the boys), or in the dolls' corner (for the girls). With three choices in each situation, a maximum of nine different children could be named by S, though a lesser number was permitted.

Physical development. Measures of Ss' physical development were based on records kept by the staff nurse. She saw all children at the center routinely each morning for a brief checkup. In addition, they all had routine dental examinations throughout the period of the study, at which times the nurse was also present. Ss were also given a thorough medical examination by the staff doctor at both the beginning and the end of the study. On the basis of these medical experiences, Ss were rated by the nurse on the following variables:

(1) The child's general health. This referred to his ability to withstand disease and excess fatigue, his ability to carry out his normal routine of activities, and his weight and body tone in relation to his own body build.

(2) The child's state of fatigue. This referred to chronic listlessness and tiredness as manifested in the morning checkups by the nurse.

(3) The child's fear reactions to the medical staff, as seen at examinations given by the doctor, dentist and nurse.

The ratings were made on 3-point scales, each variable being rated on S's condition at the end of the study, and on the change of condition from the beginning of the study.

Absence records. A record of absences was kept by the secretary of the center. She noted the number of times each S was absent and found out the reason for the absence, upon S's return to kindergarten. Reasons for S's absence were analyzed at the end of the study, and these were categorized into the following: (a) illness or visit to a hospital or doctor; (b) outings, holidays or other celebrations; (c) family mishaps, which includes no one being available to bring S to the center a given morning; (d) miscellany, which includes mainly cases where no reasons were given or where vague statements were made from which no clear reason could be ascertained.

Parents' questionnaires. To assess the parents' evaluation of how Ss were progressing at home, two questionnaires were administered, one in April and one in August. In all cases the mother filled out the form. The questions on each occasion dealt with the parent's opinions about the two types of attendance program, and about any changes noted in her child from the start of the study to the time of the questionnaire. On the first questionnaire there was an additional item, which was a request for the mother to keep a diary account for one week of the child's activities at the times he was not at the center. This was included in the questionnaire for the purpose of assessing whether the home activities differed for the two groups in spite of the precautionary matching.

Method of Data Analysis

On three of the major variables, i.e., behavioral observations, physical development and absences, the groups were compared by checking for each of the twenty pairs which member of a pair had a larger score on the particular item in question. The results of this comparison were then tested for significance by use of the binomial expansion. Since the concern in this study is only *whether* the groups differed, and not *which* group had the larger score, the test for significance is a two-tailed one. Thus, to get a significant difference at the five per cent level, fifteen out of a possible twenty members of one group would have to have larger scores than their matched partners, this giving a significance level of 4.2 per cent, the closest figure to five per cent. The probabilities were taken from Walker and Lev (*5*, p. 458, Table IV B). Ties were handled by splitting them, e.g., if the ratio is fourteen to two members of one group having larger scores than their partners, the four ties are split to give a ratio of sixteen to four, significant at the 1.2 per cent level.

In the sociometric analyses, all Ss in the groups were used, not just the twenty pairs of matched partners. This was necessary since the group structure involved all the children within the group, Ss obviously not being aware of who was and who was not in the final set of matched partners. (They were not even aware that they were so matched. All they knew was that every member of their group was observed on a given day by the investigator. Within the first few weeks they became quite oblivious even to this.) The method used to arrive at the group structure was a matrix analysis suggested by Forsyth and Katz (*1*). The manipulation of the matrix was achieved by use of the Matrix Analyzer, devised by S. B. Hammond of the Department of Psychology at Melbourne University.

As with the sociometric data, analysis of the parents' questionnaires was also not based on comparisons of the twenty matched pairs. This was necessitated by the fact that not all the parents of the forty children returned the form. The analysis was thus based on all the questionnaires returned to the investigator, which includes some Ss who were not in the matched pairs and excludes some Ss who were in the final matching. The data thus cover about eight-five per cent of all Ss in the groups. (Inspection of the data indicates that even if the fifteen per cent for whom no forms were returned were extreme deviants, the results would have been substantially the same.)

RESULTS

Behavioral Observations

It will be remembered that Ss were observed for six months, from March to August. Comparison of the two groups on this variable was made for both July and August, since inspection of the data showed that trends of change in the six-month period were not uniform. By taking two months as the terminal point from which to measure the change that occurred since the first month (from March), a more reliable criterion is established for evaluating any significant differences found in the development of the two groups.

In the analysis of over one hundred items, which made up the seventeen categories, only five items showed a significant difference between the two groups. Four of these showed a significant difference in only one of the terminal months, and one other item showed a significant change for *both* July and August, this item being "obligingly following a peer request." The nature of this significant decrease was that the S-A group had shown this item significantly more in March (at the 0.7 per cent level), and by the end of the study the S-A group had decreased in this item to the low incidence consistently displayed from March to August by the D-A group. All significant differences found between the two groups on the behavioral observations are found in Table I.

From Table I it can be seen that by the end of the study the S-A group had more parallel play and showed more coöperative acts in relation to aggressive acts than did the D-A group. But the S-A group also had more distractible behavior, and it did not keep its higher standing in the item of obligingly following peer requests. In contrast, the D-A group had made a greater improvement in relation to the S-A group by decreasing in aimless or idle preoccupation. However, although the differences were significant, except

TABLE I—SIGNIFICANT DIFFERENCES BETWEEN THE TWO GROUPS IN THE BEHAVIORAL OBSERVATIONS

Items	March	July	August
Parallel play (1d)			S-A > D-A† ($P < 0.03$)
Flitting about and/or being silly (2b)			S-A > D-A‡ ($P < 0.03$)
Σ Cooperative acts to peers/Σ aggressive acts to peers (Σ category 6/Σ category 7)		S-A > D-A† ($P < 0.01$)	
Obligingly following a peer request (8dl)	S-A > D-A† ($P < 0.01$)		
Decrease in obligingly following a peer request (from March)		S-A decr'd > D-A ($P < 0.01$)	S-A decr'd > D-A ($P < 0.03$)
Decrease in aimless or idle preoccupation (1a) (from March)*		D-A decr'd > S-A§ ($P = 0.002$)	

---

* Although the incidence of the item on any of the above three months did not differ significantly for the two groups, the groups nevertheless reversed their standing on the item. This reversal gave a significantly greater decrease from March to July for the D-A group.

† An advantage for the S-A group.

‡ A disadvantage for the S-A group.

§ An advantage for the D-A group.

for one item they did not appear in both terminal months, thus indicating a trend only and not a definitive change. Moreover, in over a hundred comparisons made for each terminal month, five items can be expected to show a significant difference by chance alone. In view of these considerations, the results must be taken to indicate that the groups did not basically differ by the end of the study, since the points of comparability significantly outweighed the points of difference.

Sociometric Data

In group structure some differences (obtained by inspection of the sociomatrices) appeared between the groups, and these differences were most apparent in the second test. On the first sociometric test (end of April) the S-A group was composed of several cliques which were fairly well integrated with one another. The D-A group in contrast had one quite large and well-integrated subgroup composed of several cliques, with a few small cliques fairly independent from this subgroup. On the whole, neither group was very well integrated, and both groups tended to show unisexual preferences. The groups were comparable in their percentage of isolates and unreciprocated choices. In degree of integration of the total group, the S-A group was slightly superior.

A marked change occurred by the second test (end of July). At this time both groups became considerably more integrated, but their method of achieving integration differed markedly. The S-A group formed two quite independent subgroups, which were composed entirely of unisexual friendships. In contrast, the D-A group expanded its earlier subgroup to include two-thirds of the members. The cohesion between these members was very strong, with much exchange of mutual choices. This large subgroup was bisexual. The rest of the group structure was made up of a small, loosely-knit cluster of the older boys, and of a few isolates. Although the D-A group had by far the greater integration, neither group achieved an ideal group structure, in that the D-A group had nearly one-third of its members peripheral to the tightly-knit core, and in that the S-A group had too little overlap between its cliques and it had a marked cleavage between the two sexes.

In spite of their differences, the two groups had some points in common. One was a marked social mobility of the members within each group, from the first to the second test. A *tau* (3) of 0.27 was obtained for the D-A group and a *tau* of 0.34 was obtained for the S-A group, these showing a marked change in the social rank of the members within each group. A further point of similarity was that with the increase of integration in the two groups, the choices made by Ss became more "realistic." To assess the validity of the choices (i.e., the degree of "reality" of the choices), the choices were compared with the names of the children with whom each S interacted during his observation period, at the time the sociometric tests were made. The mean ratio of "unrealistic" choices dropped from the first to the second test from 0.51 to 0.29 for the D-A group, and from 0.42 to 0.31 for the S-A group.

In summary, differences between the groups appeared in the type and degree of integration which each group evolved. The difference increased with time, and at the end of the study the D-A group was found to be the better integrated group. In spite of these differences, the groups were alike in showing marked social mobility and more realistic choices from the first to the second test.

Physical Development

In physical development the two groups were nearly identical. In both groups the majority of Ss (seventy-five per cent) had improved in health from March to August, though the majority (sixty per cent) did not change in their state of fatigue and in the amount of fear shown to the medical staff. In general, Ss were rated to be in good health with little or only moderate fatigue at the end of the study, in August, the majority of Ss also showing no fear reaction to the medical staff. No differential age effect was found, the overall findings adequately representing both the younger and the older Ss.

Absences

Absences were compared for all six months of the study. In total number of absences the D-A group had nearly twice as many days lost as the S-A group, the mean number of days lost being fourteen per S for the D-A group, 7.5 per S for the S-A group. Considering that the D-A program called for twice as many days of attendance, the obtained difference would be expected on an *a priori* basis. In the breakdown of "reasons for absence" the groups did not differ in any consistent way. For both groups the vast majority of absences were due to illness or visit to a hospital or doctor. The only other sizable reason for absence was outings or family vacations, and this occurred in less than twenty per cent of the total number of days lost.

Parents' Questionnaires

It will be recalled that eighty-five per cent of the parents answered the questionnaires; of these, eighty-eight per cent answered both forms. Analysis of the two questionnaires revealed that ninety-five per cent of the D-A and eighty-five per cent of the S-A parents preferred daily attendance. This strong a preference was expressed on both the first and the second questionnaires. Although the majority of the parents who answered both forms did not give the same reasons on both forms for preferring daily attendance, the types of reasons mentioned most often remained quite similar. (The questionnaires were of the open-ended type, so that the responses had to be grouped into categories after the forms were returned.) The reasons given the most number of times were: (a) the child enjoys kindergarten, in general; (b) the child enjoys the peer companionship at kindergarten; (c) home does not provide as satisfactory interests or activities; (d) kindergarten is a necessary preparation for school. These reasons all have the implicit assumption that, since kindergarten is a desirable place, there should be as much of it as possible.

On the questionnaires the mean number of positive changes per child reported on both occasions for both groups was 1.6. The two groups were quite comparable in this, on both forms. Some difference between the groups was found, however, in the percentage of children for whom positive changes were reported: for the two questionnaires, the mean per cent of Ss for whom a positive change was reported was ninety-six per cent for the D-A group and seventy-six per cent for the S-A group. Negative changes were negligible, being cited in both groups for both forms in less than ten per cent of Ss. On both forms the main positive changes reported were: (a) widened intellectual development and maturity; (b) greater confidence and social independence; (c) greater sociability with peers. The one marked difference found between the groups occurred on the second questionnaire, at which time a considerably greater number of D-A than S-A parents reported that their children had improved in the sphere of social interaction. The improvements cited by the S-A parents tended to group themselves under the area of greater intellectual maturity, which included the child's greater ability to amuse himself at home.

Data from the weekly diaries revealed that both groups had nearly identical activity experiences at the times they were at home. The vast majority of social contacts which the children had were with their immediate family, of which siblings were the most frequent contacts. The types of activities undertaken were playing (as an undifferentiated activity, i.e., the mother giving no account as to the nature of the play), going shopping with mother or similar outings, reading or being read to, and daytime sleep. Creative activities were minimal in both groups, and even "reading or being read to," which was relatively high in the activity list, was cited as occurring, on the average, only three times during a child's week.

DISCUSSION

The results indicate that the only area in which a consistent difference was discernible between the two groups was in social interaction. Both the sociometric data and the parents' reports indicated that the D-A group had increased relatively more than the S-A group in positive peer interaction. That this difference was not complete, however, was seen in the fact that the behavioral observations did not substantiate this finding. In fact, the observations showed that the groups were comparable in frequency of associative and coöperative play (terms taken from Parten (*4*), which denote a higher level of social play than is found in parallel play), and that the S-A group had more parallel play and showed more coöperative acts in relation to aggressive acts than did the D-A group at the end of the study. Both of these results would indicate superior social development of the S-A group, but they cannot be taken as reliably established since neither of these observational results occurred on both terminal months. Since the behavior observations do not substantiate the trend suggested by the sociometric data and the parents' reports, and since the observations tend to dispute this trend, only a limited conclusion can be made about differences between the groups. At best it can be concluded that though the D-A group was not observed to show more social interaction than the S-A group, when Ss in the D-A group did interact, their social play encompassed a larger group of children. It may have been this wider social scope that carried over to the home setting and prompted more parents of the D-A group than the S-A group to note positive social changes in their children.

Except for this difference between them, the two groups developed in fairly similar lines. In part this confirms the original suggestions made by Gesell, and in part it also contradicts them. Unlike the parents of the Yale studies, the parents of the present study overwhelmingly preferred daily attendance. The weekly diary also showed that during an average week the children on spaced attendance do not go on more outings than the children on daily attendance. Both these findings, which contradict the findings of the Yale group, undoubtedly are a result of the lower socioeconomic level of the parents in the present investigation. Another finding that contradicts the suggestion made by Gesell is that the children, including the three-year-olds, do not suffer from greater fatigue or ill health by daily attendance. In addition, the lack of a clear-cut superiority of one group over the other points out that assimilation of newly learned material by use of alternate days at home is not necessary. On the other hand, the present study does confirm the idea that continued attendance is also not essential for healthy progress at kindergarten.

The findings from the present study cannot be completely compared to those of the Yale studies, however, since the present groups were not made up of children of the same chronological age, as were the groups used in the Yale studies. Groups formed on the basis of a similarity in age may yield somewhat different results from those obtained with a mixed-age group as used in the present study. It is therefore possible that the original Yale findings will be borne out, i.e., that three-year-olds, in a group made up of their own age only, would profit best from kindergarten on a spaced attendance basis, while five-year-olds, in a group of their own age only, would profit best from daily attendance. One further qualification to the results may also reside in the fact that the majority of the present Ss had attended kindergarten previously. Although there is nothing in the present study to suggest it, it is nevertheless conceivable that with a group of children who have never been to kindergarten different results might have been obtained.

In view of the present results, however, it does appear that the type of attendance program, in terms of alternate or daily morning attendance, does not appreciably affect the children's general development. Thus, the type of attendance program put into operation rests on other factors, such as the needs of the community and of the individual child, as well as the parents' preference for one or the other type of program.

REFERENCES

1) E. Forsyth and L. Katz, "A Matrix Approach to the Analysis of Sociometric Data," *Sociometry*, 9: 340–347, 1946.

2) A. Gesell and F. L. Ilg, *Infant and Child in the Culture of Today*. London, Hamish Hamilton, 1943.

3) M. G. Kendall, *Rank Correlation Methods*. London, Charles Griffin, 1948.

4) M. Parten and S. N. Newhall, "Social Behavior of Pre-School Children," In R. G. Barker, J. S. Kounin, and H. F. Wright, editors, *Child Behavior and Development*. New York, McGraw-Hill, 1943.

5) Helen M. Walker and J. Lev, *Statistical Inference*. New York, Henry Holt, 1953.

PUBLICATIONS RECEIVED

ROBERT H. BECK, Editor. *The Three R's Plus: What Today's Schools Are Trying to Do—and Why.* Minneapolis: University of Minnesota Press, 1956, pp. 392. $5.00.

BENJAMIN S. BLOOM, Editor. *Taxonomy of Educational Objectives: The Classification of Educational Goals.* New York: Longmans, Green and Company, 1956, pp. 207 (paper). $1.50.

THEODORE BRAMELD. *Toward a Reconstructed Philosophy of Education.* New York: The Dryden Press, 1956, pp. 417. $4.50.

B. R. BUGELSKI. *The Psychology of Learning.* New York: Henry Holt and Company, 1956, pp. 523. $5.75.

FREDERICK I. CARPENTER. *American Literature and the Dream.* New York: The Philosophical Library, Inc., 1956, pp. 220. $4.75.

CLAY L. COCHRAN and RON ROSS. *The Case for Competition in Electric Power.* New York: Public Affairs Pamphlets, 1956, pp. 28 (paper). $0.25.

Compulsory Education in Australia. Paris: UNESCO, 1951, pp. 189 (paper). $1.00.

HAROLD F. COTTINGHAM. *Guidance in Elementary Schools.* Bloomington, Indiana: McKnight & McKnight Publishing Co., 1956, pp. 325. $5.00.

THEODOSIUS DOBZHANSKY. *The Biological Basis of Human Freedom.* New York: Columbia University Press, 1956, pp. 139. $2.95.

H. EDELSTON. *Problems of Adolescents.* New York: Philosophical Library, Inc., 1956, pp. 174. $4.75.

ROLAND C. FAUNCE. *Secondary School Administration.* New York: Harper & Brothers, 1955, pp. 387. $4.50.

ERIKA FROMM and LENORE DUMAS HARTMAN. *Intelligence: A Dynamic Approach.* Garden City, New York: Doubleday & Company, Inc., 1955, pp. 52 (paper). $0.85.

GEORGE E. GARDNER, Editor. *Case Studies in Childhood Emotional Disabilities. Volume II.* New York: American Orthopsychiatric Association, 1956, pp. 368. $5.00.

FLORENCE L. GOODENOUGH. *Exceptional Children.* New York: Appleton-Century-Crofts, Inc., 1956, pp. 428. $4.50.

EUGENE F. HAHN. *Stuttering: Significant Theories and Therapies.*

Second Edition. Prepared by Elise S. Hahn. Stanford, California: Stanford University Press, 1956, pp. 180. $4.00.

Robert J. Havighurst, Eugene Stivers and Robert F. DeHaan. *A Survey of the Education of Gifted Children.* Chicago: University of Chicago Press, 1955, pp. 114 (paper). $1.50.

Otto Hughes and Virgil E. Schooler. *A Survey of Athletics in the Secondary Schools of the North Central Association.* Bloomington, Indiana: Indiana University Bookstore, pp. 52 (paper). $1.00.

R. C. Johnson. *Psychical Research.* New York: Philosophical Library, Inc., 1956, pp. 176. $2.75.

Bruno Klopfer. *Developments in the Rorschach Technique. Volume II. Fields of Application.* Yonkers-on-Hudson: World Book Company, 1956, pp. 828. $8.50.

Henry Clay Lindgren. *Educational Psychology in the Classroom.* New York: John Wiley & Sons, Inc., 1956, pp. 521. $5.00.

Henry Clay Lindgren. *Workbook to Accompany Educational Psychology in the Classroom.* New York: John Wiley & Sons, Inc., 1956, pp. 183 (paper). $1.95.

Henry Clay Lindgren. *Instructor's Manual for Use with Text and Workbook to Accompany Educational Psychology in the Classroom.* New York: John Wiley & Sons, Inc., 1956, pp. 14 (paper). $0.50.

Margaret Mead. *New Lives for Old: Cultural Transformation—Manus, 1928–1953.* New York: William Morrow and Company, 1956, pp. 548. $6.75.

Roy Walso Miner, Editor. *Psychotherapy and Counseling.* New York: The New York Academy of Sciences, 1955, pp. 319–432 (paper). $3.50.

Thomas Munro. *Art Education: Its Philosophy and Psychology. Selected Essays.* New York: The Liberal Arts Press, 1956, pp. 387. $5.00.

M. Prudhommeau. *Les Enfants Deficients Intellectuels.* Paris: Presses Universitaires de France, 1956, pp. 305 (paper). 1.200 fr.

W. Kenneth Richmond. *Education in the U. S. A.* New York: Philosophical Library, Inc., 1956, pp. 227. $4.50.

Arnold M. Rose. *Sociology: The Study of Human Relations.* New York: Alfred A. Knopf, Inc., 1956, pp. 589. $6.00.

CHARLES SCHOENFELD. *God and Country*. New York: Philosophical Library, 1955, pp. 119. $3.00.

GEORGENE SEWARD. *Psychotherapy and Culture Conflict*. New York: The Ronald Press Company, 1956, pp. 299. $6.00.

M. BREWSTER SMITH, JEROME S. BRUNER and ROBERT W. WHITE. *Opinions and Personality*. New York: John Wiley & Sons, Inc., 1956, pp. 294. $6.00.

FRANZ STEINER. *Taboo*. New York: Philosophical Library, 15 East 40th Street, 1956, pp. 154. $4.75.

GEORGE S. STEVENSON. *Mental Health Planning for Social Action*. New York: McGraw-Hill Book Company, Inc., 1956, pp. 358. $6.50.

JEAN-M. SUTTER. *Le Mensonge chez l'Enfant*. Paris: Presses Universitaires de France, 1956, pp. 170 (paper). 480 fr.

JAMES M. THYNE. *Patterns of Error in the Addition Number Facts*. London: University of London Press, 1954, pp. 270. 10s, 6d.

THEODORE L. TORGERSON, GEORGIA S. ADAMS AND ERNEST R. WOOD. *Student's Workbook: For Use with "Measurement and Evaluation for the Elementary-School Teacher" by Torgerson-Adams*. New York: The Dryden Press, 1955 (paper). $2.00.

CHARLOTTE TOWLE. *The Learner in Education for the Professions as Seen in Education for Social Work*. Illinois: University of Chicago Press, 1954, pp. 432. $7.50.

ROBERT M. W. TRAVERS. *Educational Measurement*. New York: The Macmillan Co., 1955, pp. 420. $4.75.

FRED T. TYLER. *The Prediction of Student-teaching Success from Personality Inventories*. Berkeley: University of California Press, 1954, pp. 86. (paper) $1.25.

C. W. VALENTINE. *Parents and Children*. New York: Philosophical Library, 1955, pp. 212. $3.75.

S. J. VAN PELT. *Modern Hypnotism: Key to the Mind*. Westport, Connecticut: Associated Booksellers, 1956, pp. 88 (paper). $1.00.

J. E. WALLACE WALLIN. *Education of Mentally Handicapped Children*. New York: Harper & Bros., 1955, pp. 485. $4.50.

NORMAN F. WASHBURNE. *Interpreting Social Change in America*. Garden City, New York: Doubleday & Company, Inc., 1954, pp. 50 (paper). $0.95.

ROBERT I. WATSON. *Psychology as a Profession*. Garden City, New

York: Doubleday & Company, Inc., 1954, pp. 65 (paper). $0.95.

WILLIAM W. WATTENBERG. *The Adolescent Years.* New York: Harcourt, Brace and Company, 1955, pp. 510.

When the Mountains Move: Technical Assistance and the Changing Face of Latin America. Paris: UNESCO, 1954, pp. 69 (paper). $0.75.

DELOS D. WICKENS AND DONALD R. MEYER. *Psychology.* New York: The Dryden Press, 1955, pp. 541. $5.25.

J. E. MANCHIP WHITE. *Anthropology.* New York: Philosophical Library Inc., 1955, pp. 191. $2.75.

MARTHA WOLFENSTEIN. *Children's Humor: A Psychological Analysis.* Glencoe, Ill.: The Free Press, 1954, pp. 224. $3.75.

J. WAYNE WRIGHTSTONE, JOSEPH JUSTMAN and IRVING ROBBINS. *Evaluation in Modern Education.* New York: American Book Company, 1956, pp. 481. $5.00.

PAUL YAFFE. *It Takes All Kinds.* New York: Metropolitan School Study Council, 1955, pp. 32 (paper).

WILBUR A. YAUCH, MARTIN H. BARTELS AND EMMET MORRIS. *The Beginning Teacher.* New York: Henry Holt and Company, 1955, pp. 339. $3.50.

ROMUALD ZANIEWSKI. *Les theories de milieux et la pedagogie mesologique.* Editions Castrman, 1952, pp. 282 (paper).

EMPRESS YOUNG ZEDLER. *Listening for Speech Sounds. Stories for the Speech Clinician and the Classroom Teacher.* Garden City: Doubleday & Co., Inc., 1955, pp. 145. $3.00.

BOOK REVIEWS

DAVID ABRAHAMSEN. *Who Are The Guilty? A Study of Education and Crime.* New York, 1952, pp. 340 + XIV.

Crime increase in the United States is alarming. Between 1940 and 1950 every crime except robbery increased out of porportion to increase in population. The crime rate in the United States is higher than in Italy, which has the highest in Europe, and more than ten times as high as Scandinavia. What are the reasons for so much crime in the wealthiest country on earth? Why has it taken so many diversified forms, ranging from the singular crime such as robbery, rape, kleptomania or murder to organized form such as gambling, fire setting, insurance fraud, betting, kidnapping, climaxing in gangsterism and racketeering? The social, economic and emotional significance of crime warrants the attention of everybody, not only professional criminologists, says Dr. Abrahamsen in the question that he raises, namely, *Who Are The Guilty?* Is it society, home, school, the delinquent or the criminal himself? Everybody is involved. Not only the criminal himself. Not only the delinquent himself. Involved also are the seemingly innocent—politicians, college boys, policemen, even high officials. The statistics of crime are appalling enough to compel us to ask the question about the nature of our society. Is America really a country full of power without purpose, as it has been accused of being? And, is it our society that is guilty? Generally speaking, that's not Dr. Abrahamsen's answer—that's only a partial answer. But, so also is blaming the delinquent himself as if he were self-made—a partial answer. Who are the guilty—how are delinquents made—how can they be unmade—that is, how can they be understood and how best treated by an enlightened society—are the questions Dr. Abrahamsen considers in this interestingly written volume called *Who Are The Guilty?*

In twelve chapters, Dr. Abrahamsen considers the nature of the problem and treatment for it. Eight chapters are devoted to a consideration of the nature and extent of the problem. These include, besides general statistics, a consideration of early signs of juvenile delinquency; psychology in the criminal; psychosomatic disorders in antisocial behavior; crime and mental illness; delinquent types; sex and murder. The last four chapters are devoted to what can be considered treatment and re-education, and these are considered in chapters called Prisons and Imprisonment; Treatment and Rehabilitation; Mental Health and Schools; Education is Prevention. His attitude is expressed by him in the quote from a previous publication of his, which reads as follows: "... in treating the offender, one must keep in mind that the principle upon which the offender can be changed and educated is based upon his social experience and his reaction to it. And it is this social experience that is the basic stimulus for the individual's re-education." The appendix includes a glossary of prison jargon, a bibliography, as well as an index.

In the preface, Dr. Abrahamsen expresses the hope that this book of his becomes a primer about education in crime. It is his purpose "to make the reader see and hear so he can feel what is happening in this vast but little understood area of human behavior." To a fair extent he succeeds in doing so in this volume. He is more successful in selecting and presenting the kind of content that can yield knowledge for understanding and use that will enlighten the lay people interested in the problem. It will not, however, satisfy, for example, the people who are interested in what is the known and verified knowledge, what are the results of present research, and the strength and weaknesses of methodology used in research in this area. Too much of the material is presented in a highly personalized manner. Too frequent use is made of quotes expressed by him in his own previous publications. Generally speaking, however, the point of view is in line with modern attitudes of informed workers; familiarity with the literature is manifested by the author in spite of his more frequent reference to his own writings, and only casual reference made to studies made by other investigators. In his reaction to W. H. Sheldon's book on *Varieties of Delinquent Youth*, he is, however, more critical. What he says of it is that Sheldon stresses the constitutional factors and gives scant attention to the offender's inter-personal relationship. The facts and attitudes that he does call to the attention of the reader should serve to arouse an interest in a topic that is of importance and that does need organized and continued attention.

H. MELTZER

Psychological Service Center, St. Louis

ORIS ROSS YOST. *What You Should Know About Mental Illness.* New York: Exposition Press. 1953. pp. 165.

Not everything is known about mental health or about the treatment of mental disease, described here as "the malady of our times." But much of what is known is not applied by the people who could profit from it. The purpose of this small volume by Dr. Yost, the medical director of a sanitarium in South Carolina, is to present for the lay reader facts about mental illness, ways of preventing it, treating it, curing it. Behind the presentation is an attempt to give a more hopeful picture of the possibilities for treatment of mental illness. It is intended primarily for the individual and for his family. The material is presented in three parts. Part One, called The Mind in the Making, includes short chapters on The Rôle of The Mother, The Rôle of the Teacher, and The Rôle of the Community. Part Two, called Mental Illness and Its Treatment, describes The Sick Mind, Psychosomatic Illnesses, Techniques of Treatment, and chapters on Alcoholism, Drug Addiction, Epilepsy and Suicide. Part Three, called Mental Illness: A Public Responsibility, includes short chapters on The Care of the Mentally Ill, Mental Health: The Goal, Steps Toward Mental Health: The Community's Responsibility, and Psychiatry and Religion (with an introductory note by Norman Vincent Peale). The book contains a Foreword by Dr. Nolan D. C. Lewis, and an Introduction by Dr. Leland E. Hinsie, as well as the note by Rev. Peale.

The book is not intended to be a critical presentation of the known facts. It is, instead, presented in a manner that aims to give the reader "a feeling for psychiatry," to use Hinsie's words in his introduction. It is this feeling, Hinsie thinks, the author puts into his book that makes it attractive to the general public. It is also this feature that may receive more critical and questioning attitudes on the part of the more scientific-minded. Dr. Peale's introductory note to the last chapter on Psychiatry and Religion is perhaps better preaching than it is modern mental hygiene, when he says that "the false gods of comfort and security to which man has given his energy and loyalty must give way to a true God, the Creator and Almighty Father, under whose leadership and protection a true brotherhood may be created." This is a frame of reference which should make it possible for some psychiatrists to get along well with religio-therapists, and would probably be welcome to people who have members of their family in Dr. Yost's institution or similar institutions. If hope and faith based on popularized interpretation of knowledge is the goal of the book, then, to a fair measure, it serves its purpose. H. MELTZER

Psychological Service Center, St. Louis, Mo.

ADOLF NICHTENHAUSER, MARIE L. COLEMAN AND DAVID S. RUHE. *Films in Psychiatry, Psychology and Mental Health.* New York: Health Education Council. 1953, pp. 269.

This is an excellent reference volume for the increasingly large number of people who are in position to use films at professional meetings or meetings for popularizing their area of knowledge. The body of the book consists of 51 critical film reviews designed to help the film-user. The intent is to make it unnecessary for people who want to select films to preview them or go to the expense of spending time and money for previewing the films, before deciding concerning the suitability for intended use. Most of the reviews in the volume were prepared by Marie L. Coleman. Where this has not been the case, it is indicated and credit given. Of the 51 films reviewed, only 2 are intended for psychiatric training. These are Mental Symptoms series and Psychotherapeutic Interview series. The variety of topics included is wide, but most of them are in the field of therapy, developmental psychology or etiology of neurosis. The content is presented in three parts: Part I, the smallest, includes a gallery of scenes from the films considered, a foreword, a preface and a listing of acknowledgements. The real content begins with Part II, which contains a discussion of the film reviewing techniques and suggestions for the utilization of films in psychiatry, psychology and mental health. Included here are chapters on how to use the book, how the content and presentation influence each other in psychiatric, psychological and mental health films, the use of motion pictures for the teaching of psychiatry, and a-half-century of motion pictures in neurology, psychiatry, psychology and mental health education. Part III is composed of reviews of representative films in these three areas, and some of the topics included are: Activity for Schizophrenia, Angry Boy, Children's Emotions, Eight Infants, Family Circles, Feelings of Depression, Feeling of Hostility, Feeling of Rejection, Frustration Play Techniques, Overdependency, Problem Children. Part IV of the book is composed of indices to films and contains also an index to the content of the film subject matter of 101 films in these three fields and a suggested audience guide to the 51 films.

The appraisal in the three sections is composed of three portions: (1) the content section, which analyzes the purpose and objectives of the film and examines the treatment of the subject; (2) the presentation section, which discusses the film style used, a record of actual cases, reënactment dramatization and the skill with which the various factors of motion picture presentation have been applied; and, (3) the effectiveness section, which reaches conclusions from the analysis of content and presentation and the rest of the effects of the film. In general, it attempts to answer the questions: what can be learned from the film?, who can learn from it? and, how well can one learn from it?

All in all, there is a good deal of knowledge for use here which can help improve decisions concerning the use of films for purposes needed in teaching or popularized presentation of the many topics included.

H. MELTZER

Psychological Service Center, St. Louis, Mo.

Health and Human Relations. (A report on the third of a series of international conferences on Health and Human Relations, held in Hiddesen near Detmold in Germany, August, 1951, sponsored jointly by the Josiah Macy, Jr., Foundation and the World Federation for Mental Health.) New York: The Blakiston Company, Inc. 1953. pp. 192.

Health and Human Relations is a report of a conference, third in a series of international conferences, sponsored by the Josiah Macy, Jr., Foundation and the World Federation for Mental Health. The three co-chairmen were: Dr. Frank Fremont-Smith, Medical Director of the Macy Foundation; Dr. John R. Rees, Director, World Federation for Mental Health (who chairmanned the previous two conferences); and Dr. Werner Villinger, Director, Neuropsychiatric Clinic, University of Marburg. The editors of the English report have tried to present results of the conference in a manner that would be a projection of the experience and reflect the spirit of the meeting, and show both the organization and the dynamics of its development. Generally speaking, the editors did an excellent job in giving this kind of report. The material is presented in three parts. Part I deals mostly with the reports of the working groups, composed of groups on Mental Hygiene and Education, Mental Health in Modern Society, Sources of Social Tensions in Postwar Germany, and Promotion of a Mental Hygiene

Movement in Germany, with a list of recommendations for conference action. Part II is composed of informal presentations by select speakers, which include a discussion on Health and Human Relations, the Problem of Neurosis, The Sense of Inner Identity. Part III includes the comments by four of the participants. The volume also includes an appendix—the Conference Who's Who.

In the reports, one is informed that the purpose of the Conference is to do more than merely discuss, that many of the participants were actually doing things that have real significance for world health. Demonstrations of such influences are cited. Making decisions about the invitation list for the Conference seemed to be a problem for the people who made the preparations for it, but the list seems to be fairly representative of psychiatric and related auspices and people in the fields in Germany at the present time—a total of thirty-six participants.

The goal of this, as well as the previous two conferences, is certainly an admirable one. It is devoted to "the aim of improving health in the broadest sense of the word as a basis for truly human and dignified interpersonal relationships." Illustrations of the material considered by some of the working groups can be given by some of the questions considered by, for example, the first group on Mental Hygiene and Education. The questions considered included the meaning of health, the meaning of authority in human relationship, the meaning of freedom, the goals of education, the rôle of fear in educative process, ways of integrating religion and education, qualities of leadership needed in the world today. Some of the questions considered by the second group working on Mental Health in Modern Society were: how we can help individuals to become more independent in their thinking and thus be less influenced by mass movements; what ways of the mass media of communication, radio, film and press, affect human relationships; how we can plan for further research on the problems of mass behavior; and how we can encourage coöperative staff planning and thinking within all kinds of organizations; and, what are the qualities of leadership most needed in the world today and how can such leadership be developed. People interested in the World Federation of Mental Health doings will find this an interesting volume, including the large appendix describing in some detail the Conference Who's Who. H. MELTZER

Psychological Service Center, St. Louis, Mo.

Kenneth W. Hodgson. *The Deaf and Their Problems.* New York: Philosophical Library, 1954, $6.00, pp. 364 plus xvii.

In "The Deaf and Their Problems" the English author, Kenneth W. Hodgson, gives an excellent historical account that will be of interest to all people in the field, no matter where they live. Though the reference point for consideration of their problems is England, the knowledge considered and the history of the interest in fulfilling the social responsibility of educating the deaf is of universal nature. In this book one can find an excellent historical account of the kind and the amount of attention that was given to the education of the deaf from the sixteenth century onward. It also includes a number of chapters on the mechanism of hearing and the assessing of damage to the ear. Brought to a head also is a consideration of what has been a dilemma in methods of educating the deaf—the question of oralism (teaching by speech and lip-reading) versus silent methods (by signs and finger-spelling). Here also one can get a description of all of the important contributors to the field over the ages. Interesting by-products, also included, are the descriptions of Laura Bridgman, Helen Keller, the contribution of Alexander Graham Bell, as well as the contribution of Gallaudet.

In the preface, written by R. A. S. Paget, which includes an excellent review of the intentions of the volume, the question of definition of goals is raised—is it to educate the deaf so that they shall appear to be normal, or is it to educate their minds so that they may, as soon as possible, come out into the hearing world as its intellectual equals? That is a question for Mr. Paget. The author, himself, seems to have no dilemma of this nature and conceives of education for the deaf as having an unreserved goal, namely the achievement of the reality of full citizenship. Accordingly, the author finds the present stage of development merely the beginning of an international consideration of the problem.

He ends the book with the calling of our attention to the need for educating colored as well as white and this he does in the following words: "There are now nearly 1000 deaf schools in the world. But four out of every five of them are for white children. Tens of thousands of colored children are still in that uncouth misery to which all deaf children were once condemned. For them our four centuries of progress is something which has yet to happen."

H. MELTZER
Human Relations Research Foundation, St. Louis, Missouri

PERCIVAL S. SYMONDS. *Dynamics of Psychotherapy: The Psychology of Personality Change.* Volume I. New York: Grune & Stratton, 1956, pp. 201.

This is the first of a three volume edition. The subtitle for this volume is *Principles,* and according to the preface, "outlines the general psychodynamic principles on which contemporary psychotherapy is based." Discussion is limited to principles and does not attempt to deal with the problems of who should practice therapy and of professional relationships. The point of view is "essentially eclectic;" the purpose is "to integrate these diverse points of view into a consistent and unified whole."

In sixteen chapters the book deals with the definition, need and aims of psychotherapy; the process of psychotherapy; indications for counseling; indications for psychotherapy; contraindications for psychotherapy, the nature and definition of psychoneuroses; the dynamics of psychoneuroses, (two chapters); neurotic trends and symptoms; predisposing and precipitating factors; goals of psychotherapy (three chapters); spontaneous recovery (two chapters), and some theoretical problems including determinism and free will, psychotherapy and learning, values, culture and education.

Psychotherapy seems to have many aims: improved ability to function in work and family; ability to drop one's defenses and to face problems; inner peace and freedom from anxiety, worry and stress; self-sufficiency and self-responsibility; a help to childhood development and greater independence; honesty with self; "integrating the various parts of the mind into a smoothly running whole, so that the individual will have all of his forces under his active control;" relief from suffering; achievement of happiness; conscious recognition of unconscious goals; good adjustment; maturity; release from and reduction of tension; personality integration; improvement and strengthening of the superego; new superego standards; better insight and proper self-evaluation; a new philosophy of life; greater adequacy and competence; increased capacity for social intercourse; and even the "attainment of free will."

The author states that "Therapy fulfills its task when it helps the individual to accept himself more fully, to gain insight with regard to himself, and to find freedom in self-expression. . . . The direction that his expression will take must be determined by the forces of society." "Spontaneous recovery" is to be considered as well as recovery due to psychotherapy.

In this excellent study the reader may find some difficulties; there seem to be too many dogmatic assertions but the author excuses them because of lack of experimental data. On page twelve the thesis of the book is "that psychotherapy consists of introducing into the life of the individual some new influence which will counteract the influence of previous experiences which have given rise to his neurotic difficulties." On page nineteen the thesis is "that *abreaction* is the principle dynamic factor in therapy." It is a question as to the exact relation of the terms "psychoneurosis" and "neurosis," when they are treated separately and somewhat differently in the index.

The attempt to integrate the various points of view existing in connection with psychotherapy today may leave the reader with some doubt as to whether they can all be integrated, but he may have a high respect for the valiant effort. Psychoanalysis may appear to loom too large in this volume, but the author clearly states that it is not all of psychotherapy.

This is a valuable contribution to the study of psychotherapy. A well made index (pp. 201–211) concludes the volume.

A. S. EDWARDS

The University of Georgia

ALBERT J. HARRIS. *How To Increase Reading Ability*, Third Edition. New York: Longmans, Green and Co., Inc., 1956, $5.25, pp. 633 plus xix.

This text deserves wide use. It does little to reduce the sediment of speculation about the controversial subject of reading instruction, but neither does it muddy the pool further. The author's primary aim is to provide "scope, balance, and practicality" and, secondarily, "enough theory to provide a basis for judging when and why a particular technique may be useful." In addition, he promises "a new theory on the significance of hand and eye dominance in relation to reading disability."

He has achieved the primary aim. Topics discussed include the bulk of relevant skills, techniques and controversies, couched in simple language. Teachers will be pleased to find many useful techniques not to be found elsewhere. The author has a lucid style and a knack for approaching problem areas historically, thus providing for the student a perspective not easily achieved otherwise.

The secondary aim, that of providing theory where needed, is achieved less perfectly, at least in those instances where theory is treated explicitly, i.e., where an attempt is made to apply current theoretical constructs. The reader may be surprised to find, for instance, that "As it relates to word recognition, insight means attaining a clear and accurate perception of the printed word" (p. 319). On the other hand, when the author takes positions on controversial questions, his implicit theoretical assumptions appear to be both accurate and consistent(pp. 22, 28, 36, 80). One serious exception is that concerning the important problem of the components of reading. The weight of evidence from factor analyses of reading comprehension points to "words" and "words-in-sentences" as the only clear components. Dismissing the evidence because of apparent disagreement of "statistical experts," somewhat more justified in 1946 than in 1956, the author devotes a chapter to eight "special comprehension skills." It is questionable whether such special skills exist qua skills. They appear to be conceptual activities most easily influenced by attention to "words" and "words-in-sentences." To continue sending teachers down what appears to be a blind alley may be a serious error.

The author's new theory of hand and eye dominance in reading disability, briefly stated, consists of a common denominator for reversal symptoms, viz., directional confusion caused, perhaps, by neurological defect or slow maturation, conversion from left to right handedness, or emotional tension. The theory has considerable merit and is congruent with recent evidence from the reviewer's laboratory.

Special commendation should be given the appendices which include an annotated list of over one hundred published tests, ad-dresses of publishers and a graded book list of more than one thousand one hundred titles. This book should prove valuable both as a text and as a reference manual.

DONALD E. P. SMITH

University of Michigan

WILLIAM A. HUNT. *The Clinical Phychologist.* Springfield: Charles C Thomas, 1956, pp. 205.

This book presents the Thomas William Salmon lectures for 1954. The committee made an excellent selection for lecturer, since Dr. Hunt has not only lived through the period of the development of clinical psychology, but has been a member of committees which have had to deal with the difficult and intimate problems of relationships with other professions. It was therefore possible for him to bring to light pertinent facts that have not been generally known.

The purpose of the lectures was to bring about a better understanding of the problems and relationships, especially between psychology and medicine. The three lectures are presented under the following titles: "What Is a Clinical Psychologist?" "How the Clinical Psychologist Came to Be," and "The Problems and Future of Clinical Psychology."

The treatment is essentially historical. It begins with general psychology as the background, and proceeds to discuss the gradual development of clinical psychology upon the scientific foundations already achieved. It goes on to discuss the increasing demands and responsibilities laid upon psychologists, with special reference to the efforts to work out the proper relationships between medicine and psychiatry on the one hand, and psychology on the other. A special interest of the author is the desire to contribute to getting beyond the stage of definition and argument and back to the great and much needed task of coöperative work in promoting mental health.

As we have found historically in other fields, and as appears in connection with the problems involved in our own profession, it has been found to be more difficult to work out satisfactory definitions and limitations of duties and responsibilities, than it is to coöperate, in actual practice, with those who are willing to coöperate. Perhaps the satisfactory definitions will be possible only after continued evolution and development in the work of those who continue to practice in proper consideration with and consideration for one another. Progress in the right direction is indicated in the text.

Among the many problems discussed are diagnosis, the use of tests, the use of professional judgment, research, kinds of psychology, the nature and methods of therapy and psychotherapy, certification and licensing, psychology's claim to the three functions of diagnosis, research and therapy, the clinical psychologist's work as privately encouraged and publicly denounced, the question of who is "ill," the future of clinical psychology, etc.

This is probably the best brief, yet reasonably comprehensive statement of the work, problems and progress of clinical psychology at its present stage. It is written clearly, effectively, and with a delightful sense of humor. It is also written with good judgment in a situation in which the latter is certainly needed. The few minor adverse criticisms might be made to seem out of place in view of the general excellence of the entire volume. It is highly recommended to anyone who desires to learn accurately about clinical psychology.

The book concludes with sixty references.

A. S. EDWARDS

The University of Georgia

JAMES L. HYMES, JR. *A Child Development Point of View.* New York: Prentice-Hall, Inc., 1955, pp. 145 plus ix.

The promotion of health and the prevention of mental illness is what the author of this volume, Mr. James L. Hymes, Jr., has in mind as a goal in writing the small volume on the child development point of view. The material is presented in popularized manner, both in the selection of captions and in the write-up. The captions, themselves, pretty well describe what the author has in mind, as well as his level of presentation in the light of it. "You are a Teacher" is followed by "Your Youngsters Must Like Themselves" and "Your Youngsters Can Climb to Highest Peaks." The purpose of the volume is relatively well indicated in his last paragraph of this small book: "Once you truly regard the child as a human—caring, seeking, preferring, needing—you have to work. You have to think, to use your judgment and your sensitivity. But the motor is running, and the course is charted. How far will the human go? How fast? For how long? With your good treatment, new records are possible: of reason and decency, of kindliness, of caring. The task is well worth the effort of your hands, the time of your head, and the searching of your heart."

This little volume is not intended as a text in educational psychology. It can serve excellently as a public relations manual for the inspiration of school teachers.

H. MELTZER
Human Relations Research Foundation, St. Louis, Missouri

CHARLOTTE TOWLE. *The Learner in Education for the Professions.* Chicago: The University of Chicago Press, $7.50, pp. 432.

The Learner in Education for the Professions as Seen in Education for Social Work—the full title of this volume—precisely describes the nature of the volume, in content as well as in interpretation. Charlotte Towle is an experienced social worker, who has played an active rôle and has been a part of the growth of psychiatric social work and talks from direct observation and experience. What she says is of particular importance to social workers, but much of what she says is also of significance to professional workers interested in education in their own areas.

The material is presented under two parts. Part I is called "Learning Process and Educational Process" and considers objectives of professional education, basic learning principles and also personality development and the integration of learning. This part also considers administrative dimensions such as selection of students and the social component in education for professions.

In Part II, called "Applications in Teaching Social Case Work" are included a consideration of the place of social case work, areas of learning in case work, selection and arrangement of case material, faculty collaboration and the content of class sessions, assignments and examinations.

The general attitude of the author is given in the early part of the first chapter: "Professions today are intent upon their contribution to the development of the individual rather than with services that merely assure survival. It is fast becoming the common sense that, to serve mankind for its development, one must understand man. . . Responsible, intelligent, professional help must increasingly be oriented to the individual in the context of his social position, with perception of what the service is doing to the person while doing something for him."

The book should be of interest to all people concerned with professional education, particularly in the area of the social sciences. It includes a bibliography, as well as an index.

H. MELTZER
Human Relations Research Foundation, St. Louis, Missouri

GEORGENE SEWARD. *Psychotherapy and Culture Conflict.* New York: The Ronald Press Company, 1956, pp. 299.

The author speaks of this volume as a collaboration: he supplied the textual material and Dr. Judd Marmor supplied the four case studies from his own practice. The purpose of the book "is to delineate various cultural aspects of psychotherapy that have hitherto been slighted. . . . The major portion of the book is devoted to an exposition of the psychodynamics peculiar to members of a variety of subcultures in the United States today."

Taking "classical psychoanalysis" as the point of departure, the author examines neo-Freudian systems, and attempts to "point the way a step further, across the cultural gaps separating various minority groups from the mainstream of the culture."

Eleven chapters take up in the following order, A Preface to Psychotherapy, Old World Accents, The New Look, Social Class and Personality, Treatment and Status, Negro Portraits, Color and Conflict, Women as a "Minority," New World Symphony and Discord, America's First Families, and Jews Between Two Worlds.

It appears that the study of subcultures is now one of the most important problems on the frontiers of psychological investigation and that psychotherapy is in need of therapists well acquainted at least with the subculture involved in any given case. For any treatment it is believed that a proper understanding of cultural effects is necessary.

The author does not accept psycho-analytical hypotheses uncritically and warns against interpreting in terms of wrong theories. Emphasis is placed upon concrete examples; definite contribution to the understanding of the general discussions is made by the case studies of Dr. Marmor.

These case studies include (1) Flight from Poverty, dealing with a thirty-three year old attorney; (2) Color and the Blues, a study of a twenty-seven year old Negro saxophone player; (3) A Female Woman Hater, a married woman in her early thirties; and (4) A Jewish "Fascist," bringing out factors involved the fascist trends of a young American college student.

The main discussions also include more or less detail relating to the cultural problems of the Hopi, Navajo and Sioux Indians; Greeks; Japanese; Italians; Chinese; Issei and Nisei; and a few others who are included in the very complex matrix of cultures in The United States. Space is especially accorded to problems of anxiety, authority, childhood and child training, dependency, ethnic identification, family problems, hostility, learning and education, sex problems, social class and mobility, values, and therapy.

Whether or not the reader is inclined towards psychoanalytic interpretations, he will find in this book valuable suggestions for better insight into problems of personality with special reference to the influence of cultures and the conflicts between cultures. Examples given indicate how completely psychotherapy can fail when the facts of cultural conflicts are not known and understood.

Each chapter contains a list of pertinent references. There is an index of names (pp. 285–289) and an index of subjects (pp. 291–299).

A. S. Edwards

The University of Georgia

THE JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 48 October, 1957 Number 6

THE LEARNING OF MOTOR SKILLS AS INFLUENCED BY KNOWLEDGE OF MECHANICAL PRINCIPLES

Frances M. Colville

Santa Barbara College, University of California

Many general principles of mechanics are relevant to the teaching, learning, and performance of activities included in the physical education curriculum. These principles may describe the motion of objects such as balls and racquets, movement of the body itself, or a combination of both.

The question has been raised as to whether knowledge of these general principles and an understanding of their application to these activities will facilitate the learning and improve the performance of pupils in physical education classes. Also of interest are questions concerning the influence of knowledge of these principles as applied to one activity upon subsequent learning of other activities to which the same principles apply. Results of a few experiments suggest that a pupil who understands a principle related to one skill may master a related skill more readily than the pupil whose experience has been restricted to specific instruction in technique without explanation of pertinent principles.

With reference to the teaching of a motor skill in which a specific principle of mechanics is involved, then, two questions have been raised:

1. What is the effect of knowledge of a principle upon immediate learning of a skill to which the principle applies?
2. What is the effect of knowledge of a principle learned in relation to one skill upon subsequent learning of a different or more complicated skill to which the same principle is applicable?

Reported investigations bearing on the acquisition and transfer of principles are limited, and although most authorities agree that

Bureau Ednl. Researc

general principles are probably transferable, little strong experimental evidence is available to support such a claim. Apparently, little or no experimentation has been done in the field of physical education on the effect of knowledge of principles of mechanics in learning situations involving body movement and large muscle activities.

Judd (**7**), and Hendrickson and Schroeder (**4**) reported that knowledge of the principle of refraction was beneficial in dart throwing and rifle shooting at a target under water. Cox (**2**) found that knowledge of principles involved in manipulating electrical equipment facilitated the assembling and stripping of such equipment. Other experimenters have reported similar results from investigations involving mathematics, mental games, spelling, card tricks, and mechanical puzzles (**13, 11, 15, 10, 14, 8, 2**). In contrast to these findings are a number of investigations in which the method of including instruction in the principles involved did not appear to be more effective. In studying the effect of drill versus the learning of generalizations related to addition and subtraction, Olander (**9**) found that the two methods appeared to be equally effective. Babitz and Keys (**1**) found similar results with college chemistry classes, and Hendrix (**6**) reported only one difference, significant at the 12% level, in her experiment with teaching algebraic generalizations.

With the exception of the experiments by Judd, Hendrickson and Schroeder, and Cox, all of the reported investigations have dealt with various types of learning other than motor learning, and only Judd's experiment involved skills requiring the coordinated use of large muscles of the body as well as fine coordinations of the fingers. Judd, and Hendrickson and Schroeder, while reporting a difference in favor of instruction in principles, included this instruction in such a way that additional time was allotted to the group receiving the instruction. Thus, the experimental groups had the advantage of additional time as well as additional information. Furthermore, Judd supplied no statistical analysis of his early experiment, and the other differences reported were significant at a statistical level below that which is usually considered acceptable.

In the present investigation, an attempt was made to control these ambiguities by using an experimental design which insured that both the experimental and the control groups spent equal amounts of time in learning and performing and by considering only those differences which were significant at the 5% level of confidence or above.

Method

Purpose of the Study

It was the purpose of this study to investigate certain questions related to the teaching of physical education activities in which specific principles of mechanics are involved. This problem was approached by: (*a*) selecting three principles of mechanics which are pertinent to motor skills; (*b*) selecting three motor skills, each of which utilizes one of the principles; (*c*) establishing for each skill two comparable groups of *S*s, one of which was taught without reference to the principle involved, and the other of which was taught to understand and apply the principle; (*d*) comparing for each skill the performance of the two groups, one of which spent the entire time practicing the skill without reference to the principle involved, and the other of which spent part of the time practicing the skill and part of the time in learning the principle; (*e*) comparing for each skill the performance of the two groups in a similar or more complicated form of the skill to which the same general principle applies.

Selected Principles and Skills

The selected principles and the skills in which they are utilized were as follows:

Principle I: The angle of incidence is approximately equal to the angle of reflection. Skill: Rolling a ball against a surface, or surfaces, from which it would rebound.

Principle II: In stopping a moving object, the force opposing the momentum must be equal to the force of the momentum, and if the object is to be caught, this momentum must be dissipated by reducing the resistance of the catching surface. Skill: Catching a tennis ball in a lacrosse stick and catching a badminton bird on a tennis racquet.

Principle III: An object set in forward motion through the air by an external force is acted upon by the force of the momentum and by gravital acceleration. Skill: Archery.

Procedure

Three parallel experiments were devised, each of which was designed to investigate performance of one of the three skills. The *S*s for all three experiments were undergraduate women students at the University of Southern California. There were 36 *S*s in the Ball Rolling experiment, 40 in the Catching experiment, and 42 in the Archery experiment. (The *S*s in the Archery experiment included some men whose scores were treated separately from those of the women.) For each of the experiments the *S*s were divided into two groups, one of which spent the entire time learning and practicing the skill, and the other of which spent part of the time learning about the principle and the rest of the time learning and practicing the skill. Each experiment consisted of two tests. During the First test the *S*s of the nonprinciple group learned the skill and practiced. The *S*s of the principle group learned the principle, and the skill, and practiced. The total amount of time spent by both groups was the same. During the Second test both groups practiced a similar or more complicated skill. Each group spent the same amount of time and no mention was made of the principle involved.

The Ball Rolling experiment. The Ball Rolling experiment was designed to illustrate the principle describing the rebound of a ball. The appratus consisted of a modified pinball machine plunger, a squash ball, a felt-covered table surface four feet wide and six feet long, and a target made of colored construction paper. The overall size of the target and each of its scoring divisions was determined by recording the hits made by a group of comparable *S*s who were not used in the experiment. These *S*s rolled the ball so that it would rebound from one side of the table as close as possible to the center of the adjacent side which was marked with a square of white paper. The rest of the side was covered with brown wrapping paper, and spots on this paper to which the balls rolled were marked in red pencil. This scoring sheet provided the basis for establishing a six-standard deviation scale, and from this scale the scoring areas of the target were derived.

The *S*s were asked during the First test to roll the ball against one side of the table in such a manner that it would rebound to the target placed against the adjacent side. For the Second test they were asked to roll the ball so that it would rebound from two adjacent sides to the target set at right angles on the third side.

The Catching experiment. The Catching experiment consisted of two tests each involving the skill of catching a moving object. The First test involved catching a tennis ball in a lacrosse stick. For the Second test, they caught a badminton bird on a tennis racquet. The apparatus for this experiment, in addition to balls, birds, crosse, and racquet, included a homemade device for projecting the tennis ball and the badminton bird. This device proved to have high reliability.

The Archery experiment. The Archery experiment was planned to illustrate the principle of gravital acceleration and its action on arrow flight and the relation of line of flight to line of sight. The First test involved shooting from 20 yards. The Second test included ends shot from 30 yards, from 40 yards, and a Junior Columbia Round.

Analysis of data. In each experiment, the reliability of the scores was established by use of the split-half technique, the odd scores being correlated with the even scores. Use of the Spearman-Brown Prophecy Formula for predicting reliability of the whole test resulted in coefficients which ranged from .81 to .94 and thus it was felt that these scores were sufficiently reliable to permit further analysis and comparison.

All of the data were analyzed by means of analysis of variance using Edwards' (**3**) technique for repeated measurements of the same subjects. This technique tests differences between methods, between trials, and between the interaction of trials and methods. In each experiment the initial ability of the two groups, as evidenced by preliminary trials, was tested and found not to differ significantly.

Findings

In the Ball Rolling experiment, no significant difference in the performance levels of the two groups was found on either test. A difference between trials significant at the 1% level of confidence was found on both tests.

In the Catching experiment, no significant differences were found in performance levels, and a difference between trials significant at the 5% level of confidence was found only on the First test.

In the Archery experiment no significant difference in performance levels was found. A difference significant at the 1% level for men and at the 5% level for girls was found on the Second test at all distances except forty yards.

The following general findings were noted:

1. A significant amount of learning took place under both methods of instruction.

2. This learning was not only similar in amount but also in pattern. Exceptions occurred in the Second test of the Catching experiment where no significant increase in scores occurred in either group, and in the Junior Columbia Round for men in the Archery experiment, where both groups increased in scoring ability, but the group which had been taught the principle was significantly better at twenty yards.

These findings are in general agreement with those of some published investigations and are opposed to those of some others. However, most of the reported experiments have dealt with various types of learning other than motor learning. The ability to use understood principles in performing a motor skill presents a complex problem since more than one type of learning is involved. As Ragsdale (**12**) has pointed out, a learner may understand a principle when applied to inanimate objects, but he may not understand and be able to apply it to his own movements. In general, it seems that the findings of this study support Ragsdale's observation, at least in the initial stages of acquiring an unfamiliar motor skill.

Conclusions

Within the limitations of the three parallel experiments which constitute the present investigation, there is no evidence:

1. That instruction concerning mechanical principles utilized in the performance of a motor skill facilitates the initial learning of the skill to any greater extent than an equivalent amount of time spent in practicing the skill.

2. That such knowledge facilitates subsequent learning as evidenced in the performance of a similar or more complicated skill to which the same principle is applicable.

However, since it appears that some part of the learning period may be devoted to instruction concerning general principles without detriment to the motor learning of the students, it would seem desirable to include such instruction in order to provide this additional opportunity for acquiring some related knowledge about principles of mechanics and the application of forces.

References

1. Babitz, M., & Keys, N. Experiment in teaching pupils to apply scientific principles. *Sci. Educ.*, 1939, **23**, 367–370.
2. Cox, J. W. Some experiments on formal training in the acquisition of skill. *Brit. J. Psychol.*, 1933, **24**, 67–87.
3. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
4. Hendrickson, G., & Schroeder, W. H. Transfer of training in learning to hit a submerged target. *J. Educ. Psychol.*, 1941, **32**, 205–213.
5. Hendrickson, G. & Schroeder, W. H. Transfer of training. In R. G. Kuhlen & G. G. Thompson (Eds.), *Psychological studies of human development.* New York: Appleton-Century-Crofts, 1952.
6. Hendrix, Gertrude. A new clue to transfer of training. *Elem. Sch. J.*, 1947, **48**, 197–208.
7. Judd, C. H. Special training and general intelligence. *Educ. Rev.*, 1908, **36**, 38–42.
8. Katona, G. *Organizing and memorizing.* New York: Columbia Univer. Press, 1940.
9. Olander, H. T. Transfer of learning in simple addition and subtraction. *Elem. Sch. J.*, 1931, **32**, 358–369; 427–437.
10. Overman, J. R. Experimental study of the effect of the method of instruction on transfer of training in arithmetic. *Elem. Sch. J.*, 1930, **31**, 183–190.
11. Peterson, J. C. The higher mental processes in learning. *Psychol. Monogr.*, 1920, **28**, No. 7 (Whole No. 129).
12. Ragsdale, C. E. How children learn the motor types of activities. In N. B. Henry (Ed.) *Yearb. nat. Soc. Stud. Educ.*, 49, Part I. *Learning and Instruction.* Chicago: Univer. of Chicago Press, 1950.
13. Ruger, H. A. The psychology of efficiency. *Arch. Psychol.*, 1910, **2**, No. 15.
14. Thiele, C. L. The contribution of generalization to the learning of addition facts. *Teach. Coll. Contr. Educ.*, 1938, No. 763.
15. Waters, R. H. The influence of tuition upon ideational learning. *Psychol. Monogr.*, 1920, **28**, No. 7 (Whole No. 129).

Received February 6, 1957.

STUDENT VS. INSTRUCTOR DESIGN OF STUDY METHOD[1, 2]

SLATER E. NEWMAN[3]

Air Force Personnel and Training Research Center, Colorado

The assigning of study tasks to be done by the student either in or out of class is common educational practice. Such practice, however, implies an assumption that each student can best prescribe for himself how such material should be studied.

It was the purpose of the present investigation to demonstrate that such an assumption is not warranted; that rather, the use of study materials and procedures designed on the basis of information provided by research on learning will lead to faster student learning than the use by each student of his own set of techniques.

Procedure

The task confronting each student was that of learning the names for each of 20 electrical symbols. This task was chosen because it appeared to be representative of the kind of task on which considerable research-based information already exists. The following schedule was used with each *S*, all of whom were run individually:

1. A pretest was administered to determine how many of the 20 symbols the student could name correctly. The papers of those students who knew the names of more than three symbols were not used in the analysis of post-test scores.

---

[1] This report is based on work done under ARDC Project No. 7715, Task No. 77283, in support of the research and development program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government. The opinions or conclusions expressed herein are those of the author. They are not to be construed as necessarily reflecting the views or endorsement of the Department of the Air Force or of the Air Research and Development Command.

[2] A paper, based on some of the findings reported here, was presented at the 1956 meetings of the American Psychological Association.

[3] Now with the Department of Psychology, North Carolina State College, Raleigh, North Carolina.

2. Instructions dealing with what the student should do during the learning period were read aloud by the instructor.
3. Each student studied the list either for 7¾ or for 12 minutes. It was anticipated that such differences in study time would produce statistically significant differences in performance on the posttest and that these differences might be used as a rough yardstick in gauging the effects of the kind of procedure used during the study session.
4. The posttest (which was the same as the pretest) was then administered. All 20 symbols were presented on a sheet of paper, and the student was asked to write the name of each symbol he knew to the right of the symbol. The posttest was administered immediately after the study period.

Students were randomly assigned either to a Student Group or to an Instructor Group. The procedures used during the study period with each group were the following:

Student Group ($N = 15$). At the beginning of the study period, the student was given a sheet on which were listed each of the 20 electrical symbols and its name, and he was asked to study the list. Each name was to the right of its symbol. The order in which these symbols were arranged on the study sheet differed from that in which they appeared on the test sheet. Seven students were allowed 7¾ minutes for study, and eight students were allowed 12 minutes for study. The instructor was present during the entire study period.

Instructor Group ($N = 15$). Suggestions for techniques to be used with this group were obtained from five psychologists familiar with the results of research on learning the kind of task employed here. To each psychologist the symbol-name pairs to be learned were presented, and he was asked to suggest techniques which could be expected to facilitate the learning of these pairs. There was not, however, complete agreement among the five psychologists as to techniques which should be employed. Thus, the combination of techniques used with the Instructor Group was based on a selection by the author from among those suggested. For the final procedure, file cards were prepared on which appeared either the symbol and its name, the symbol alone, the name alone, or the name of the set (to be described next) to which a symbol belonged. The 20 symbols were divided into four sets of five each. All symbols in a set shared one feature in common. For set one, it was a circle; for set two, a wavy line; for set three, a black dot; and for set four, a straight line. Early in training, study of set one was completed before study of set two was begun, etc. Approximately 10 seconds intervened between study of each set. These materials were used in a five-step procedure as follows:

Step 1. Each of the five names was presented individually for approximately five seconds (e.g., capacitor). The instructor pronounced each name and then had the student pronounce the name.

Step 2. The set name was then presented (e.g., straight lines). For this and subsequent steps, a three-second rate of presentation was employed for each card.

Step 3. Each of the five symbol-name cards was then presented (e.g., capacitor symbol-capacitor). The student pronounced each name and was corrected if he erred.

Step 4. Each card with the symbol alone was then presented (e.g., capacitor symbol), and the student responded with the name of the symbol if he felt he knew the name.

Step 5. After these steps had been followed with each of the four sets, each of the 20 symbols was presented alone and was followed by a card on which appeared the symbol and its name.

With seven students, this set of procedures was used once (total study time for each student approximately 7¾ minutes). With eight additional students, this set of procedures, excluding Step 1 (i.e., response learning) was then repeated a second time. The total study time for this group was approximately 12 minutes.

Subjects

Thirty airmen, each of whom had recently completed basic training, were randomly assigned to one of the experimental groups. All *S*s had a score of 5 or above on the Airman Classification Battery's Technical Specialty Aptitude Index (a measure of general intellectual ability).

Results

The random assignment of students to groups did not result in significant between-group differences in aptitude. Nevertheless, each student's score on the posttest was adjusted on the basis of his aptitude index. A 2 × 2 analysis of variance (i.e., type of study × amount of study time) of these adjusted scores was then accomplished and is presented in Table 1. The *F*s for both type of study and amount of study were found to be significant at the .01 level. The mean scores for each cell of the design are presented in Table 2.

The significant *F* for type of study indicates that there were

TABLE 1

ANALYSIS OF VARIANCE OF ADJUSTED SCORES

Source	*df*	Mean Square	*F*	*p*
Type of study	1	10,046.70	10.51	.01
Amount of study time	1	12,080.38	12.64	.01
Interaction	1	234.00	.24	
Within	26	955.50		
Total	29			

TABLE 2

MEAN APTITUDE AND RETENTION TEST SCORES FOR EACH GROUP

Group	Amt. of Study Time	*N*	Mean Aptitude Index	Mean Retention Score (Unadjusted)	*SD*[a]
Instructor	7¾ min.	7	7.4	9.6	5.33
Student	7¾ min.	7	6.7	13.1	1.59
Instructor	12 min.	8	6.8	13.5	2.91
Student	12 min.	8	7.1	17.0	2.00

[a] Bartlett's chi square for these data is 7.36 ($.10 > p > .05$).

real differences between the average performance scores of the Student and the Instructor Groups. The differences, however, were not in the direction anticipated. Instead, students who studied using their own techniques did better on the posttest than students for whom study techniques were prescribed.

A second study was then conducted to determine whether the introduction of certain changes in the procedure used with the Instructor Group would result in faster learning by *S*s of this group than by *S*s who studied using their own set of techniques. Two changes in procedure were introduced with the Instructor Group: (*a*) prompting training was used throughout (i.e., immediately following presentation of a stimulus-response pair, the stimulus-term was presented alone, *S* being required to respond with its response-term); thus, *S* was never tested during training, and (*b*) the items were presented in a fixed order, both within sets and between sets. This revised set of procedures was used with 26 *S*s. A Student Group ($N = 26$) received the same treatment as the Student Group in the first study. The *S*s in each group were allowed 7¾ minutes for study. A median test applied to the posttest scores of both groups resulted in a chi square of 7.69 ($p < .01$), the performance of the Student Group again exceeding that of the Instructor Group.

Discussion

An examination of the differences in procedures used with the Student and the Instructor Groups might shed some light on these performance differences. Of particular interest is the fact that for members of the Student Groups, all symbols and names were present throughout the study period. The implications here are that (*a*) this probably allowed for more opportunity to observe differences between symbols and between names than was the case with the Instructor Groups; and (*b*) students could spend more time studying those symbols and names with which they were having difficulty.

A second difference between the groups is that members of the Student Groups were probably more familiar with the techniques which were used (and which each student selected) during the learning period than were students of the Instructor Groups. Thus, the posttest performance differences between the Student and Instructor groups may not have been so great had students of the Instructor Groups had some practice in learning how to learn from the techniques used during their study period.

Finally, the manner in which the techniques were combined for use with the Instructor Groups may have led to slower learning (e.g., the relatively little emphasis on response learning).

At least two conclusions *might* appear warranted on the basis of the results of these studies: (*a*) that allowing the student to decide on the techniques he will use during study on tasks like that of learning the names for electrical symbols is good educational practice, and (*b*) that in respect to the design of materials and procedures for learning on such tasks, the experimental psychology of learning has little to offer. Such conclusions do not appear to present a complete picture.

The situation, rather, may be described like this: The effective design of materials and procedures for tasks even so apparently simple as learning the names of electrical symbols is not easy. Much information is already available from laboratory research on learning which might prove useful in designing materials for such tasks. However, until considerable experience is accumulated about how such information can best be applied to the design of materials and procedures for learning in classroom type situations (or until additional research-based information is available), the practice of allowing the student to decide on the techniques he will employ during study of the kind of task used here may be a defensible one.

Received December 10, 1956.

RETROACTIVE INHIBITION AND FACILITATION IN THE LEARNING OF SCHOOL MATERIALS

DAVID P. AUSUBEL, LILLIAN CUKIER ROBBINS, AND ELIAS BLAKE, JR.

University of Illinois

This study is generally concerned with some of the major determining factors and processes involved in the retention of typical school materials. More specifically, it deals with (*a*) the relative importance of proactive and retroactive inhibition in the forgetting of such materials, (*b*) the role of similarity between original and interpolated tasks in retroactive inhibition, and (*c*) the relative facilitating effects of simple repetition and explicit retroactive enhancement of the discriminability of learning material on its subsequent retention.

The paradigm of retroactive inhibition is congruent with the highly paradoxical "interference" theory of forgetting. According to this theory, the introduction of *identical* material subsequent to learning results in consolidation and an increment in performance but the more *similar* the subsequently introduced material is to the original learning task, short of complete identity, the more it supposedly interferes with the retention of the latter. The resulting decrement in performance (as well as the interfering processes responsible for the decrement) is designated as "retroactive inhibition."

Practically all of the experimental work on retroactive inhibition has been conducted with materials (e.g., nonsense syllables, lists of paired adjectives) that are quite far removed from the type of learning that occurs in the classroom. This approach has not been completely without rationale. First, nonsense syllables, as additive units of equal difficulty, make for greater quantifiability. Second, because they are relatively meaningless, their use makes it possible to rule out the indeterminable influence of meaningful antecedent conditions that obviously vary greatly from one individual to another. Unfortunately, however, although the use of nonsense syllables adds undoubted methodological rigor to the study of retroactive inhibition, their very nature limits the applicability of findings to the retention of meaningless materials—or to a type of learning that is rare both in everyday situations and in the school room. Nevertheless, despite the absense of any convincing a priori grounds for believing that retroactive inhibition occurs in the same way for meaningful as for relatively meaningless (arbitrary, rote, discrete) learning, findings from the latter kind of situation have been commonly and uncritically extrapolated to the former.

What reasons are there for suspecting that retroactive inhibition might operate differently for meaningful than for meaningless materials? First, nonsense syllables are discrete and isolated entities that are almost completely unrelated to the vast reservoir of concepts and information that a pupil brings with him to a learning situation. They are relatively unaffected by previous learnings; and, hence, any forgetting that results from the studying of subsequently introduced materials (over and above the forgetting of an equated group that "rests" between initial test and retest) can be plausibly attributed to the interfering influence of these latter materials. This is simply not the case in school learning. Almost any new meaningful material is related in some way to what has been learned in the past; in fact, the curriculum is deliberately organized in this fashion to provide for the gradual and untraumatic introduction of new facts and concepts. Hence, it seems reasonable to suppose that any school learning will involve interaction with established knowledge in the learner, and that this interaction will probably have greater influence on retention than the immediate interactive effects of material introduced subsequent to such learning. Furthermore, the retroactively introduced materials also have significant connections with the residue of prior learning experience. Thus, whatever effect they do have on retention is probably less a reflection of interaction with the learning task in their own right than of the activation and mobilization of relevant subsuming concepts in the individual's existing cognitive structure.

Second, in dealing with meaningless material that cannot be related to established concepts in cognitive organization, a simple paradigm of retroactive interference, e.g., stimulus or response generalization, can adequately explain forgetting. Meaningful materials, however, are invariably related to an existing cognitive structure that is hierarchically organized in terms of highly stable and inclusive conceptual clusters under which are subsumed less stable and more specific illustrative data. As they enter the cognitive field, they are appropriately "catalogued" under relevant conceptual foci, since this is the most orderly and efficient way of organizing large quantities of material for ready availability. At first, for a variable period of time, the recently catalogued data can be dissociated from their subsuming concepts and are reproducible as individually identifiable entities. Because it is most economical and least burdensome, however, to subsume as much material under as inclusive categories as possible, the import of many specific illustrative items is assimilated by the generalized meaning of the more established and highly conceptualized subsuming foci. When this happens, these items are no longer dissociable or identifiable as entities in their own right and are said to be "forgotten." Hence, barring repetition or some other special reason, e.g., primacy, uniqueness, affective content, for the perpetuation of dissociability strength, specific items of experience that are supportive of an existing conceptual entity tend eventually to be irreversibly subsumed by the generality of the latter. In other words, it is proposed that subsumability under established concepts, rather than retroactive interference due to similarity per se, is the determining factor in the forgetting of meaningful learning material.

Third, similarity between original and interpolated materials presumably bears a different relationship to retroactive inhibition when the materials involved consist of conceptual data and related information than when they consist of rote, nonsense tasks. In the latter instance, specific differences in content, in the absence of any transferable conceptual meaning, completely obviate the possibility of retroactive facilitation and only generate confusion. It is little wonder, then, that degree of similarity (short of identity) between original learning task and interpolated material is positively related to degree of retroactive inhibition. Meaningful materials, however, in addition to being characterized by specific detail and context, have a more general conceptual content that is transferable to comparable learning situations. Hence, contextually dissimilar but substantively similar interpolated learnir can retroactively facilitate the retention of previously learned materials.

A final theoretical issue that we shall explore in this study has to do with the mechanism of retroactive facilitation in the retention of meaningful material. By retroactive facilitation is meant the effect of interpolated material in *reducing* the retention loss caused by proactive inhibition. It is demonstrated when, after interpolated learning, significantly less loss occurs from immediate test of original retention to later retest than occurs when no learning task is interpolated between test and retest.[1] Consideration of the "subsumption-dissociability" theory of forgetting proposed above suggests that a given ideational element is irreversibly subsumed under an established concept not because of similarity between them per se, but because its distinguishing features are not sufficiently salient to enable it to persist as a discriminably distinct entity in its own right. Thus, it is conceivable that identical or substantive repetition of the learning material induces retroactive facilitation by increasing its discriminability from already established concepts. Another possible way of retroactively enhancing the discriminability of learning material is explicitly to point out similarities and differences between it and potentially subsuming concepts. Since this latter method presumably results in more explicit appreciation of the distinctive properties of the learning material, it should lead to greater retroactive facilitation than an equivalent amount of simple repetition.

Hypotheses

Proceeding from the foregoing theoretical considerations, therefore, an experiment was designed to put to empirical test the following hypotheses regarding the retention of meaningful school materials: (*a*) Proactive inhibition, i.e., irreversible subsumption of learning material by established concepts, accounts for more forgetting than does retroactive inhibition. (*b*) The interpolation of substantively similar (but not identical) material between initial test and retest results in retroactive facilitation rather than retroactive inhibition. (*c*) Less retroactive facilitation occurs when the interpolated material merely recapitulates the learning passage than when it explicitly delineates similarities and differences between the latter and presumed subsuming concepts.

Method

Subjects

An experimental group and three control groups were constituted out of a total population of 188 undergraduates in eight sections of a course in educational psychology at the University of Illinois. The mean ages of these groups varied from 21.8 to 22.4. The experiment was rationalized as a laboratory exercise in the psychology of learning and was performed during regularly scheduled class hours. In order to maximize ego involvement, the *S*s were informed that after the data were processed their individual scores, as well as the class results, would be reported to them.

---

[1] If sufficient retroactive facilitation occurs between test and retest, as when the learning task is repeated frequently during the interpolated sessions, an actual increment rather than a reduced decrement results.

Learning Material and Test of Retention

The learning material used in this study was a passage of approximately 1700 words on the history, sacred literature, doctrine, and ethical teachings of Buddhism. It was specially designed for use in an experimental learning situation that would bring into play the same kinds of cognitive variables and interactional processes that affect the long-term retention of typical ideational material learned in a school setting. Hence, it had to satisfy the major criterion of being largely unfamiliar in its specific aspects but still generally related to a body of knowledge relatively well established in most of our *S*s. On purely a priori grounds it seemed reasonable to assume that the facts and ideas contained in this passage would interact with and be subsumed under existing concepts of Christianity, and that the latter concepts would furnish the basis for whatever proactive and retroactive inhibition might occur in the course of learning and retention.

A 37-item multiple-choice test was constructed to test immediate and later retention of the Buddhism material. Preliminary tryouts of the passage and the test indicated that the material was actually unfamiliar but learnable, and that the test was reliable and structurally sound. Scores on the test showed a satisfactory range of variability and were distributed normally. The corrected odd-even coefficient of equivalence was .81. When the test was administered to naive *S*s who had not previously studied the passage, the mean score (number of correct items) was only 8.9 as compared to a chance expectancy score of 7.4 (one-fifth of 37).

Because of the limited duration of class periods, the passage could not be made long enough to provide a sufficient number of conceptual items for two reasonably adequate tests. Practical considerations, therefore, made it necessary to use the same test as a measure of both immediate and later retention. Hence, it is conceivable that memory of answers given initially competed in some instances on the retest with the *S*s' recollections of the material itself. The probability of this happening, however, was minimized by the eight-day interval between tests and the difficulty of specifically remembering one of five highly similar and complexly worded alternatives. In any event, this factor operated in similar ways for the experimental and control groups.

Experimental Procedure

The *S*s in all groups were first instructed to read the Buddhism passage at their normal reading speed (eight to ten minutes) and then to utilize the remainder of the 25 available minutes to commit the concepts and facts in the passage to memory. It was explained that a multiple-choice test would be given immediately after the reading was completed and also eight days later. Both tests were given as announced.

Twenty-four hours after the initial test, 20-minute sessions of interpolated learning were introduced as follows: The Experimental Group studied a 2100 word comparative essay on Buddhism and Christianity. This essay encompassed all of the material included in the Buddhism passage, but similarities and differences between the two religions were explicitly pointed out. Control Group A restudied the Buddhism passage. Control Group B studied a 1700 word passage on the history, sacred literature, doctrine, and ethical teachings of Christianity. In each instance it was explained that knowledge of the material studied during the interpolated session would be tested for a week later by the previously announced retest. Control Group C was not subjected to an interpolated learning session.

Results and Interpretation

Means and standard deviations of all groups on the initial test and retest are shown in Table 1. Since raw scores on this multiple-choice test include correct answers attributable to guessing, incidental knowledge, and the presence of unintentional cues in the five alternative answers, corrected means were computed by subtracting from the raw means the mean score of *S*s who took the test without studying the Buddhism passage. These corrected means were utilized in making all intragroup and intergroup comparisons of test scores. The distribution of test scores did not diverge significantly from the normal. *F* tests for homogeneity of

TABLE 1

MEANS AND STANDARD DEVIATIONS OF IMMEDIATE AND LATER TESTS OF RETENTION FOR EXPERIMENTAL AND CONTROL GROUPS

			Test I			Test II			Percentage Retained[b]
Group	Interpolated Material	*N*	Raw Mean	Corr. Mean[a]	*SD*	Raw Mean	Corr.[a] Mean	*SD*	(Test II/ Test I)
Experimental	Comparative	43	24.4	15.5	5.47	25.6	16.7	5.25	107.7
Control A	Identical	50	22.5	13.6	6.14	23.2	14.3	5.75	105.1
Control B	Subsuming	44	24.7	15.8	4.79	21.9	13.0	4.83	82.3
Control C	None	51	24.5	15.6	5.84	22.0	13.1	6.13	84.0

[a] Corrected means were computed by subtracting 8.9 (the mean score of *Ss* who took the test without studying the learning material) from the raw means.

[b] Corrected means were utilized in computing the percentage retained from test to retest.

variance between groups indicated that we were dealing with comparable groups (i.e., drawn from the same population).

The relatively high degree of retention (84 %) demonstrated by Control Group C after an eight-day interval without any intervening repetition can be attributed to several factors. First, the gradient of forgetting is known to be very gradual when meaningful material is memorized. Second, the availability of memory traces is maximized when a recognition rather than a recall method of retention is employed. Third, a type of reminiscence effect was partially operative as frequently becomes evident when immediate retention of partially learned, highly conceptualized material is compared to retention several days later. Dissociability of newly subsumed materials presumably tends to be at a minimal level immediately after the initial shock of interaction with their subsuming concepts, and improves spontaneously for a limited time thereafter as the temporary confusion generated by the interaction gradually dissipates. Thus, although the eight-day interval in this instance undoubtedly exceeded the temporal limits of the spontaneous recovery phase, the ratio of later to immediate retention was probably still inflated by the temporary inhibition of retention immediately after the completion of learning.

Proactive versus Retroactive Inhibition

To test the hypothesis that most forgetting of meaningful material is reflective of proactive rather than retroactive inhibition, the retest scores and percentages of retention (Test II/Test I) of control groups B and C were compared (see Table 1). Control Group B had studied an interpolated passage on Christianity, i.e., the material potentially most capable of inducing retroactive inhibition in relation to the Buddhism learning passage. Control Group C had not been exposed to an interpolated learning session. Thus, in the first instance, both retroactive and proactive inhibition were operative, whereas in the second instance only proactive inhibition was operative. The data show conclusively that subsequent study of relevant subsuming material results in only negligibly more forgetting of the original learning material than when the latter material interacts spontaneously with established subsuming concepts that the individual brings with him to the learning situation. Since most ideational material that is learned in a school setting is related to and hence potentially subsumable by an existing body of stable concepts in the learner's cognitive organization, this finding is generally applicable to school learning. It suggests that retroactive inhibition is an important determining factor in forgetting only in artificial laboratory situations employing nonsense learning tasks, and is at most a relatively minor variable in the retention of most newly learned school materials.

The Role of Similarity in Retroactive Inhibition

That similarity short of identity between learning and interpolated tasks does not cause retroactive inhibition in the case of meaningful material, but may or may not induce retroactive facilitation, is evident from intergroup comparisons of the retention data of the Experimental Group and Control Groups B and C (see Table 1). This type of analysis supports the view that the influence of interpolated learning on retention is a function of whether or not it increases the discriminability of the original learning material and thereby counteracts irreversible subsumption of the latter by established concepts. As already pointed out, although interpolation of similar subsuming material (Christianity passage) apparently does not induce retroactive inhibition, neither does it reduce proactive inhibition. A t test between the retest scores (or the percentages of retention) of the Experimental Group and Control Group C, on the other hand, shows that the interpolation of similar comparative material results in a highly significant amount of retroactive facilitation ($p < .001$). Despite the seven-day interval between interpolation and retest, slightly more (7.7 %) material was remembered on Test II than on Test I. The net conceptual residue of the comparative passage, which consisted of recapitulated Buddhism material, subsuming Christianity material, and an explicit comparison between the two religions, unequivocally increased the discriminability of the original Buddhism material. It appears, therefore, that in dealing with meaningful school learning, the absence of point-for-point-identity between original and interpolated learning tasks not only fails to induce retroactive inhibition but is also compatible with the occurrence of retroactive facilitation as long as the interpolated material enhances the discriminability of the learning task.

Repitition versus Explicit Comparison

To what can the facilitating effects of the interpolated comparative passage be attributed? One possible explanation, emphasized in our theoretical discussion above, is that explicit delineation of the similarities and differences between the learning passage (Buddhism) and its presumed subsuming concepts (Christianity) enhanced the discriminability of the former and made it more resistive to irreversible subsumption by the latter. However, our experimental design had to take into account the obvious fact that a comparative presentation must of necessity recapitulate in a different context the very same ground covered by the original learning material. It is patently impossible to compare Buddhism to Christianity without discussing Buddhism. Hence, it could be plausibly argued that the retroactive facilitation following interpolation of the comparative passage was largely induced by the recapitulative features of the latter material; and since there was, in fact, no significant difference between the Experimental Group (comparative interpolation) and Control Group A (repetitive interpolation) in the ratio of later to immediate retention scores (see Table 1), this argument appears incontrovertible. It must be concluded, therefore, (*a*) that most if not all of the retroactive facilitation associated with the interpolation of the comparative passage can be attributed to substantive repetition of the learning passage in another context, and (*b*) that *retroactive* use of comparative material is no more efficacious in counteracting proactive inhibition than simple recapitulation of the learning material itself. These conclusions, of course, do not detract from the finding that, where meaningful learning is involved, interpolation of nonidentical material not only does not necessarily result in retroactive inhibition but may also induce retroactive facilitation.

The failure of the comparative approach to reduce proactive inhibition any more than was accomplished by simple repetition can be attributed perhaps to the fact that this method may only be effective in *preventing* irreversible subsumption caused by *initial* lack of discriminability. Once material, by virtue of such initial limitations in discriminability, can no longer be dissociated from its subsuming concepts at the time of original learning, repetition may be the only possible way of retroactively enhancing discriminability. This explanation is being tested presently by means of a research design which contrasts the retention of *S*s initially learning the comparative passage with the retention of matched *S*s initially learning the Buddhism passage.

Summary and Conclusions

Retroactive inhibition and facilitation in the retention of meaningful school materials was studied in a population of 188 undergraduates divided into an experimental and three control groups. All *S*s studied a 1700-word passage on Buddhism and were tested by means of a multiple-choice examination for retention of same immediately after learning and eight days later. Interpolated learning sessions were instituted 24 hours after original learning and were arranged as follows: The Experimental Group learned a passage explicitly comparing Buddhism and Christianity. One control group restudied the Buddhism passage, and another control group studied a passage dealing solely with Christianity. The remaining control group was not subjected to any interpolated learning.

Intergroup comparisons of the ratios between initial and later retention supported the hypotheses that, where meaningful school material is involved, (*a*) proactive rather than retroactive inhibition is the determining factor in forgetting; and (*b*) interpolated learning of material substantively similar to but not identical with the original learning task does not result in retroactive inhibition, but induces as much retroactive facilitation as identical repetition of the learning material. Explicit comparison of the learning material and its presumed subsuming concepts was no more effective than simple repetition of the learning material in inducing retroactive facilitation. The findings were interpreted in the light of a "subsumption-dissociability" theory of retention and forgetting.

Received December 17, 1956.

THE IDENTIFICATION OF NUMBER ANXIETY IN A COLLEGE POPULATION

RALPH MASON DREGER

Jacksonville University

LEWIS R. AIKEN, JR.

Emory University

Many persons report in clinical sessions and in academic classes that they are emotionally disturbed in the presence of mathematics. It has been the writers' experience that reactions of this nature to mathematical and arithmetical materials constitute a major field of emotionality in academic situations. And yet, although many studies have been conducted to determine optimum ways of presenting arithmetic materials (**7**), almost no controlled research has been attempted in the realm of emotional problems associated with arithmetic and mathematics. Schonell (**12**) observed that backwardness in arithmetic is due as much to emotional as to intellectual factors, and psychoanalytic writers (**6, 8, 11**) suggest that failure in arithmetic may be related to maternal overprotection.

Inasmuch as there has not been any attempt to determine if actually there is a set of reactions to numbers and mathematics on the scale that appeared evident to the writers, an inquiry was called for to investigate the extent and nature of such reactions. Consequently, the study reported here is an endeavor to detect the presence of a syndrome of emotional reactions to arithmetic and mathematics, tentatively designated "Number Anxiety." We speak of a syndrome, for anxiety is probably more accurately described as "anxieties" rather than "anxiety" (**2**). The existence of specific situational anxieties has been shown in several investigations (**3, 13**). Moreover, O'Connor's (**10**) finding of five distinct factors on the Taylor Scale of Manifest Anxiety shows that several dimensions of anxiety are measured by this popular scale. Number Anxiety, too, may turn out to be a complex variable.

Hypotheses

Three hypotheses were formulated on the basis of previous observation concerning the existence of Number Anxiety and its relations to concomitant human functioning:

1. Number Anxiety exists as a factor distinct from (although it may accompany) "general anxiety."

2. Number Anxiety is not related to intelligence as measured by the Wechsler-Bellevue or ACE scales, although it is probably negatively correlated with ACE Q scores.

3. Persons with high Number Anxiety will tend to make lower scores in mathematics than others, even though intelligence (as measured above) between these two groups is not significantly different.

Subjects

Seven hundred and four students in basic mathematics classes at Florida State University.

Procedure

The Taylor Manifest Anxiety Scale (**14**), omitting three items of relatively low validity and including three items (Table 1) specifically designed to measure feelings of anxiety concerning working with numbers, was administered to the 704 students at the close of class periods during which mathematics tests had been returned to them. The *S*s were informed that the results of the Taylor Scale, called the "Confidential Data Sheet," would be used for scientific purposes and would have no influence on their grades or careers.

Scores on the three Number Anxiety items were correlated separately and together with the total score on the 47-item Taylor scale. In addition, intercorrelations were obtained among the three Number Anxiety items and the 14 items on the Taylor scale shown in two previous studies (**1, 5**) to be the most valid, making a 17 × 17 correlation matrix. This matrix was subjected to a Holzinger and Harman cluster analysis (**4**).

According to total scores on the three-item Number Anxiety items and the 47-item Taylor Scale, the 704 *S*s were divided into four groups, those high in both "general anxiety" and Number Anxiety, those high in one and low on the other and vice versa, and those low in both. From each of these four groups ten *S*s were chosen by a table of random numbers for further investigation. Inasmuch as males and females were not significantly different in scores on either scale, the sex variable was randomized. The 40 students in the subsample were not informed of the basis of selection, and were presumably ignorant of the connection between the administration of the "Confidential Data Sheet" and the following procedures. The *S*s in the subsample were given Form I of the Wechsler-Bellevue Intelligence Scale (**15**) by one examiner. At the same time as the administration of the verbal subtests, pre- and poststimulus readings were taken on a psychogalvanometer registering palmar skin resistance (**9**). Besides the recording of GSR deflections for each subtest item, the deflection was noted for the arithmetic subtest instructions, "I want to see how good you are in arithmetic."

The final Mathematics 105 grades for the 704 students and their American Council on Education Psychological Examination scores, Q, L, and Total, were obtained and correlated with Number Anxiety scores. Correlations were also found comparing grades, ACE, and Wechsler-Bellevue Verbal, Performance, and Full Scale IQ's with Number Anxiety scores for the 40 *S*s investigated more intensively.

Results

Number Anxiety as defined by the three scale items (Table 1) proved to be present in approximately one-third of the 704 students; 239 or 35 % responded positively to two or three of the key items.

With some minor differences, scores on the 47-item Taylor scale formed a distribution very similar to that of Taylor's original score distribution. Like the original distribution, the one obtained in this study is positively skewed; it was assumed, however, that it probably represents a normal distribution sufficiently well to allow correlations which employ a dichotomy of scores. Correla-

TABLE 1

BISERIAL CORRELATIONS BETWEEN THREE "NUMBER ANXIETY" ITEMS AND TOTAL SCORE ON TAYLOR SCALE SPLIT AT MEAN

Item	Correlation with Taylor Scale
3. I am often nervous when I have to do arithmetic.	.32
9. Many times when I see a math problem I just "freeze up."	.37
38. I was never as good in math as in other subjects.	.09
Total, items 3, 9, 38 combined	.33

TABLE 2

FACTOR LOADINGS IN "LITTLE FACTOR ANALYSIS" OF THREE "NUMBER ANXIETY" ITEMS

Item	Factors A	 B
3	−.002	.552
9	.266	.324
38	.518	.001

tions between the Taylor scale and the Number Anxiety items, individually and *in toto*, are presented in Table 1.

The relatively high correlations among the Number Anxiety items (Table 3) led the writers to suppose that they represent a cluster or syndrome; but the fact that they are not any more highly intercorrelated further suggested that the three items are not factorially pure. With the full recognition that a factor analysis of three items is purely an exploratory measure, a factor analysis, which was termed a "little factor analysis" to distinguish it from a serious one, was performed employing only the Number Anxiety items 3, 9, and 38, by means of Thurstone's centroid method. Two factors accounting for all of the common factor variance emerged. Orthogonal rotation allowed for no psychological meaningfulness, but oblique rotation to simple structure (as far as the latter can be obtained with three items) yielded greater significance (see Table 2). Factor A was identified very tentatively as "Negative math. reaction" and Factor B as "Nervousness in the presence of math."

A cluster analysis (**4**) of the correlation matrix (Table 3) of 14 Taylor Scale and three Number Anxiety items revealed three distinct clusters, tentatively labelled: I. Compulsive inattentiveness (composed of items 6 and 41 from the Taylor Scale); II. Covert, conscious strain (items 47, 45, 42, 48, 40, 44, 21, 5, 27, 29, 49, and 34); and III. Nervousness and difficulty in arithmetic (items 3, 9, and 38). The second cluster approaches "general anxiety" and the third clearly represents what is called here Number Anxiety.

Log conductance changes of the GSR (**9**) subjected to analysis of variance revealed that only on the arithmetic subtest and arithmetic instructions were there significant F ratios (at the .05

TABLE 3

TETRACHORIC CORRELATIONS AMONG 17 ITEMS FROM CONFIDENTIAL DATA SHEET

	5	6	9	21	27	29	34	38	40	41	42	44	45	47	48	49
3	.41	.10	.67	.37	.33	.24	.12	.50	.20	.03	.14	.28	.32	.35	.31	.10
5		.44	.37	.70	.59	.29	.37	.09	.52	.30	.42	.52	.67	.50	.54	.23
6			.14	.26	.28	.32	.31	.02	.40	.82	.40	.31	.50	.51	.38	.31
9				.33	.36	.16	.22	.64	.21	.18	.17	.06	.24	.19	.24	.12
21					.37	.29	.53	.27	.13	.31	.40	.50	.71	.34	.60	.27
27						.43	.38	.59	.45	.29	.56	.59	.62	.55	.55	.33
29							.16	.08	.30	.26	.32	.47	.47	.38	.28	.28
34								−.09	.46	.37	.36	.41	.31	.39	.29	.10
38									−.01	−.08	−.03	.01	−.02	.14	−.01	.07
40										.48	.47	.35	.48	.57	.68	.39
41											.48	.30	.48	.52	.47	.36
42												.47	.50	.58	.45	.38
44													.58	.37	.51	.13
45														.58	.52	.37
47															.55	.55
48																.49

level). That is, on Information, Comprehension, Digit Span, and Similarities there were no significant differences between Number Anxiety and "general anxiety" individuals or between high and low scorers on each anxiety scale. None of the interactions was significant, even on the two arithmetic comparisons. A breakdown of the arithmetic and arithmetic instructions GSR changes showed that all of the significant differences in each case lay between high and low scorers on Number Anxiety.

Correlations between Number Anxiety and final grades in Mathematics 105 for both the 40 subsample group and the 704

TABLE 4

BISERIAL CORRELATIONS BETWEEN "NUMBER ANXIETY" AND FINAL GRADES IN MATH, WECHSLER-BELLEVUE IQs, AND ACE SCORES

No. of Subjects	Grades	Wechsler-Bellevue			ACE		
		PS	VS	FS	Q	L	T
40	−.51	.20	−.015	.44	−.25	−.08	−.04
704	−.44						

original sample, together with the correlations of Wechsler-Bellevue and ACE tests with Number Anxiety, are presented in Table 4. The only correlations which are significant are those between grades and Number Anxiety.

Discussion

Several lines of evidence lead to the position that probably there is an overall dimension that may be termed Number Anxiety which differs from "general anxiety." The low correlations of the Taylor scale scores with the individual Number Anxiety items as well as with the total of the latter is the first piece of evidence. The second is the result of the cluster analysis which sharply separates Number Anxiety from the two "general anxiety" clusters. The third line of evidence is the highly significant changes which occurred in galvanic skin resistance for "number anxious" persons on the arithmetic instructions and arithmetic subtest of the Wechsler, changes which did not occur for the "general anxiety" group.

Although the identification of this overall dimension seems fairly secure, the actual unidimensionality of Number Anxiety is not assured by the study at hand. In fact, there is reason to believe that there are at least two subfactors in the surface factor isolated here, the two brought to light by the "little factor analysis" of Number Anxiety items. A further analysis of the three latter items in the matrix of 17 items is in progress. Also, an analysis of further items designed to ferret out Number Anxiety is being conducted with the 40 *S*s who constituted the original subsample.

Nothing in the present study yields bases for determining the etiology of Number Anxiety. From previous thought and spontaneous remarks by *S*s of the subsample, a number of hypotheses have been formulated as to the origin of this type of anxiety which *S*s find very handicapping. And very tentatively in the continuing research we are putting some of these hypotheses to the test. However, more work is needed to identify just exactly what Number Anxiety is and what differentiates it from other kinds of anxiety before much exact research can be done in determining etiology. One hypothesis which is a partially etiological one, that Number Anxiety is merely the result of low intelligence, seems to be set aside by the research already done, in that neither the Wechsler nor ACE showed any significant relation to Number

Anxiety. That actual accomplishment in mathematics is associated with lack of Number Anxiety is demonstrated by the present research, though how much causation is involved is a question, for one may cause the other, vice versa, or both may be the resultant of other factors.

One caution must be voiced in connection with the changes in GSR and Number Anxiety. These variables are associated, but the GSR cannot be strictly considered to be a measure of Number Anxiety, for other conditions of stress have also been shown to have GSR correlates. Apparently the Wechsler Verbal Scale is not sufficiently stressful to generate significant GSR changes except in connection with "number anxious" persons and Arithmetic.

Conclusions

All three of the proposed hypotheses were substantially supported by the research described here.

1. Number Anxiety does appear to be a separate (composite) factor from "general anxiety," although the correlation of .33 between the Taylor Scale and the three Number Anxiety items indicates that some relation probably holds between these two gross syndromes.

2. Number Anxiety does not seem to be related to general intelligence as measured by either the Wechsler-Bellevue or ACE scales. The supposition that Q scores on the ACE are correlated negatively with Number Anxiety must wait further research, for the r of $-.25$ does not reach significance.

3. Persons with high Number Anxiety do tend to make lower mathematics grades, as indicated by the r's of $-.44$ for the 704 *S*s and $-.51$ for the 40 subsample group between Number Anxiety and grades in Mathematics 105.

References

1. Buss, A. A. A follow-up item analysis of the Taylor Manifest Anxiety Scale. *J. clin. Psychol.*, 1955, **11**, 409–410.
2. Cattell, R. B. The chief invariant psychological and psycho-physical functional unities found by P-technique. *J. clin. Psychol.*, 1955, **11**, 319–343.
3. Gordon, E., & Sarason, S. B. The relationship between "test anxiety" and other anxieties. *J. Pers.* 1955, **23**, 317–323.

4. Holzinger, K. Jr., & Harman, H. H. *Factor Analysis, a synthesis of factorial methods*. Chicago: Univer. of Chicago Press, 1941.
5. Hoyt, D. P., & Magoon, T. M. A validation study of the Taylor Manifest Anxiety Scale. *J. clin. Psychol.*, 1954, **10**, 357–361.
6. Klein, E. Psychoanalytic aspects of school problems. In A. Freud, H. Hartmann, & E. Kris (Eds.), *Psychoanalytic study of the child*. Vol. III–IV. New York: International Universities Press, 1949. Pp. 369–390.
7. Kramer, G. A. *The effect of certain factors in the verbal arithmetic problem upon children's success in the solution*. Baltimore: Johns Hopkins Press, 1933.
8. Levy, D. M. *Maternal overprotection*. New York: Columbia Univer. Press, 1947.
9. Nichols, R. C., & Daroge, T. An electronic circuit for the measurement of the Galvanic Skin Response. *Amer. J. Psychol.*, 1955, **58**, 455–462.
10. O'Connor, J. P., Lorr, M., & Stafford, J. W. Some patterns of manifest anxiety. *J. clin. Psychol.*, 1956, **12**, 160–163.
11. Plank, Emma N., & Plank, R. Emotional components in arithmetical learning as seen through autobiographies. In A. Freud, H. Hartmann, & E. Kris (Eds.), *Psychoanalytic study of the child*. Vol. IX. New York: International Universities Press, 1952.
12. Schonell, F. J. *Diagnosis of individual difficulties in arithmetic*. Edinburgh: Oliver and Boyd, 1937.
13. Silverman, R. E., & Blitz, B. Learning and two kinds of anxiety. *J. abnorm. soc. Psychol.*, 1956, **52**, 301–303.
14. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
15. Wechsler, D. *The measurement of adult intelligence*. Baltimore: Williams and Wilkins, 1944.

Received February 13, 1957.

THE RELATION OF DEPARTURE TIME AND RETENTION TO ACADEMIC ACHIEVEMENT

ABRAM M. BARCH

Michigan State University[1]

For over three decades there has been a slight but constant interest in the study of persistence and its relationship to achievement, especially academic achievement. Despite this interest, several promising suggestions and studies have not been followed up. Thornton (**5**) hypothesized that the more the test of persistence approximates the school situation, the higher would be the correlation between persistence and school success. Briggs and Johnson (**2**) found that a persistence measure obtained in the achievement situation itself (namely, time to complete an achievement test given without a time limit) was related to achievement when general verbal ability was taken into account. French (**3**) developed a short test of persistence which he found resulted in an improvement in prediction of academic success above that obtained by conventional aptitude tests. His test consisted of 30 number series problems given with the instructions that some problems were insoluble and were to be so marked and was scored in terms of amount of time spent on the test.

Problem

The present study had two objectives. The first objective was to repeat the Briggs-Johnson study in a situation where more information would be available about the aptitudes of the students and the characteristics of the achievement test items and to see if a better estimate of the predictive value of their persistence measure could not be obtained. The second objective was to determine the effect on achievement test score of requiring all students to remain in the testing situation for a period of time distinctly longer than most students would stay without restraint.

The assumption behind the second objective was that the merely spending more time in the testing situation, irrespective

---

[1] The author wishes to thank the Board of Examiners of Michigan State University for the assistance rendered in carrying out this study.

of the reason(s) for spending this additional time, might be the factor responsible for increased achievement. The particular method selected for retaining students in the testing situation beyond their usual departure time was chosen to interfere least with the normal procedure of achievement testing. In addition to this aspect of minimal interference, the reason given for the detention was chosen so as to be irrevelant (or neutral) to any given person's level of persistence or achievement aspiration level.

Procedure

Achievement test

At Michigan State University, all students are required to take a one-year sequence in natural science unless they have had courses in that area at other institutions or unless they can pass special qualifying examinations in the area. The third term of this course is concerned with the atomic-molecular theory and the related concept of energy. It includes laboratory exercises on gas laws, atomic theory and electricity, and the nature of matter.

The final examination of the course, new each term, is prepared by a group of course instructors working in conjunction with members of the Board of Examiners. The test consists of 120 multiple-choice five-alternative items. The test items are intended to measure comprehension rather than knowledge of specifics. Classification of test items in the examination used in this study (spring of 1955) according to categories based on the *Taxonomy of Educational Objectives* (**1**) indicated that the test contained 8 items on knowledge of specifics, 38 items on knowledge of ways and means of dealing with specifics, 45 items concerned with ability to make use of materials or ideas through interpretation, 14 items concerned with the use of abstractions in particular and concrete situations, and 15 items concerned with the analysis of relationships.[2]

Procedure

The final examination was given to all students in the course at the same time in the same building. Instructional classes were grouped for examination purposes. The number of students per room when tested varied but averaged about 60.

---

[2] Clarence Nelson performed the classifications.

Ednl.

Subjects were students in two sets of six classrooms each so selected that no student was in an examination room with a proctor who was also his natural science instructor. In all rooms distribution of the tests was supposed to begin at 10:10 A.M.

In six of the rooms there was no change from normal testing procedure. The proctors merely recorded the departure time of students in units of five minutes each from the time the tests were distributed. The first student left just less than an hour after the test began. The testing session was terminated at 12:30 or shortly thereafter. The students in these rooms formed the initial basic group.

In the other six testing rooms, the proctors were requested to write the following on the blackboard: Please remain seated until 12:00. This is necessary to reduce disturbance by those finishing early.

The proctors were also asked to call the students' attention to the notice before beginning the test administration. Three of the instructors failed to follow the full instructions and the students in these classrooms were omitted from further analysis. The remaining students formed the initial experimental group.

Analysis

In order to obtain a more homogeneous group with respect to educational background, only those students who had entered the University in fall, 1955, and who had taken the entire orientation test battery were retained for further study. This group included a small proportion of transfer students in addition to freshmen. The remaining basic groups were composed of 190 men and 108 women. The remaining experimental groups were composed of 134 men and 48 women. Men and women were treated separately because intercorrelations between tests and achievement measures had been found to be related to sex at M.S.U. with the women usually showing higher intercorrelations than men.

The aptitude tests in the orientation test battery that were utilized in this study included the L and Q portions of the ACE (1940 version), a locally prepared Reading Test (RT) and a locally prepared Arithmetic Proficiency Test (A). The intercorrelations of these tests and the predictive value of the RT and the A for a large group of M.S.U. freshmen has previously been published (4). The RT was the best single predictor of first term grade point average.

Results

Significance of departure time

Inspection of the correlational plots relating departure time to the other measures indicated that the relationships could be assumed to be linear. This is in contrast to the results obtained by Briggs-Johnson who generally found a U-shaped trend between departure time and achievement. In the present study, the longer the students stayed the better was their achievement score.

Table 1 gives the product moment intercorrelations for natural science final (NS), departure time (DT), RT, A, L and Q. Values for men appear above the diagonal; values for women are below the diagonal. The relationship between DT and NS is low. This was expected since verbal ability was expected to obscure the relationship between these two measures; i.e., leaving at a given time after the beginning of the test means one thing when done by a person with good verbal ability and something else when done by a person with low verbal ability.

It will be noted that DT correlated negatively with all other measures except NS and A for both men and women. The positive relationship with A and the tendency for A and DT to include similar variance in multiple combinations was somewhat puzzling. Further investigation suggested the apparent reason. A was given with such a liberal time limit as to be essentially an untimed test, and the scores on this test apparently included some variance due to this untimed aspect.

TABLE 1

INTERCORRELATIONS OF ORIENTATION TEST SCORES, DEPARTURE TIME AND NATURAL SCIENCE FINAL SCORE[a]

	1	2	3	4	5	6
1. Natural Science Final		.16	.24	.43	.35	.23
2. Departure Time	.14		−.25	.10	−.18	−.16
3. M.S.U. Reading Test	.44	−.22		.25	.63	.29
4. M.S.U. Arithmetic Test	.51	.13	.35		.27	.45
5. L scale, ACE	.50	−.10	.77	.38		.44
6. Q scale, ACE	.32	−.15	.47	.68	.45	

[a] Men (190) above diagonal; women (108) below diagonal.

TABLE 2

PARTIAL CORRELATION COEFFICIENTS BETWEEN NATURAL SCIENCE FINAL SCORE AND DEPARTURE TIME

Comparisons	Men (190)	Women (108)
$r^{a}_{12.3}$	.23**	.27**
$r_{12.4}$	.13	.09
$r_{12.34}$	.18*	.20*
$r_{12.5}$	.24**	.22*
$r_{12.6}$	.20**	.20*
$r_{12.56}$	.25**	.24*

* .05 level of significance.

** .01 level of significance.

[a] See Table I for names of variables.

Table 2 gives the first and second order partial correlation coefficients between the various predictor measures and NS. All partials are significantly different from zero at the .05 level or better except the first order partials with A as the partialled variable.

Another way of evaluating the importance of departure time as a measure for predicting academic achievement is to determine the extent to which it increases the magnitude of a multiple correlation. Table 3 lists the R's obtained for combinations of the predictor variables with and without DT when NS and grade point average for first M.S.U. year were the criterion variables.

Some loss in cases occurred in the group with grade point average as the criterion. Excluded from this group were students who did not return to school the next fall term and students who had less than 32 credits or more than 63 credits at the end of their first M.S.U. year. The remaining groups were composed of 146 men and 90 women.

In all comparisons except one, addition of DT to the predictor variables significantly increased the size of R at the .05 level or better. The one exception which involved the prediction of NS for women using RT and A as predictors was borderline to significance at the .05 level. The increase in predictive variance in all cases was small, but it must be remembered that departure time is essentially one response. The results suggest the value of obtaining additional measures of this type for use in prediction of academic achievement.

TABLE 3

MULTIPLE CORRELATIONS PREDICTING NATURAL SCIENCE FINAL SCORE OR FIRST YEAR GRADE POINT AVERAGE WITH DEPARTURE TIME EXCLUDED AND INCLUDED

Criterion Variable	Predictor Variables	R's for Men	R's for Women
Natural Science	3[a], 4	.45	.58
Final Score	2, 3, 4	.48	.60
(Men = 190)	5, 6	.36	.51
(Women = 108)	2, 5, 6	.43	.55
First Year Grade	3, 4	.36	.73
Point Average	2, 3, 4	.40	.77
(Men = 146)	5, 6	.39	.53
(Women = 90)	2, 5, 6	.45	.57

[a] See Table 1 for names of variables.

Sex differences

The mean NS score for men (basic group) was 61.4 and that for women (basic group) was 56.5. The difference was significant at the .01 level. Men and women were not significantly different in DT, RT, A or L. The men scored higher (.01 level) on the Q scale than did the women.

The point of interest in these results is that the sexes did not differ in terms of test persistence. It is possible that this finding is due to the nature of the test and that somewhat different results would obtain in a situation where the women would be superior in test achievement.

Effect of Retention

The results of the basic groups indicated that 74% of the women and 85% of the men could have been expected to leave by 12:00, the end of the retention period for the experimental groups. An analysis of covariance was performed for men and women separately using RT score as the control variable. The F in both comparisons failed to reach the .05 level. Thus, no evidence was obtained in support of the hypothesis that merely spending more time in the testing situation would lead to better achievement. It should be recalled that the manner of retention was designed to provide minimal interference with normal testing procedure and the reason given for the retention was neutral with respect to a given person's level of persistence or aspiration level. The next question to be asked is whether instructions stressing the importance of taking more time on the test will affect achievement.

Conclusions

1. Persistence as measured by time taken by college students to complete an achievement test was found to be significantly related to test achievement and grade point average when verbal ability was partialled out.

2. The persistence measure significantly improved multiple regression predictions for academic achievement.

3. No difference was found between men and women in test persistence.

4. Retaining students in the testing situation with instructions which were neutral with respect to achievement had no affect on achievement.

References

1. Bloom, B. S. (Ed.) *Taxonomy of educational objectives*. New York: Longmans, Green, 1955.
2. Briggs, Arvella, & Johnson, D. M. A note on the relation between persistence and achievement on the final examination. *J. educ. Psychol.*, 1942, **33**, 623–627.
3. French, J. W. The validity of a persistence test. *Psychometrika*, 1948, **13**, 271–277.
4. Jackson, R. A. Prediction of academic success of college freshmen. *J educ. Psychol.*, 1955, **46**, 296–301.
5. Thornton, G. R. The use of tests of persistence in the prediction of scholastic achievement. *J. educ. Psychol.*, 1941, **32**, 266–273.

Received February 26, 1957.

DIFFERENTIATING CHARACTERISTICS OF A GROUP OF STUDENTS HAVING PSYCHOLOGICAL PROBLEMS

George Douglas Mayo

Naval Air Technical Training Command, Memphis[1]

Virtually everyone associated with an educational program has had occasion to observe students with psychological problems and, frequently, to inquire whether or not the school situation is in some degree responsible for the condition observed in the student. Rotter (**5**, pp. 416–423) has discussed this matter in an astute manner and has attempted to treat it within the general framework of his social learning theory. In addition to the orientation described in the following paragraphs, a section of the present paper has been devoted to consideration of the findings of the study in terms of Hull's system (**3**) as adapted to emotional behavior by Brown and Farber (**1**).

It is important from the practical point of view to know whether psychological problems tend to be more closely associated with factors in the immediate environment or more closely associated with factors more remote from the current situation. If it is known that maladjustment is associated primarily with factors in the immediate environment, it may be possible to change these factors in such a way as to alleviate the condition. If not, it is necessary to look to other, probably more difficult, remedial procedures.

The study was conducted within this general framework of situational versus nonsituational causation of psychological problems and disorders. But no claim can be made that it provides definitive results concerning this general problem. In this connection it should be noted that it was necessary to use available variables which could be classified as situational or nonsituational instead of following the more desirable procedure of sampling a universe of situational and nonsituational variables. The study does, however, identify and cross validate certain variables which differentiate to a statistically significant extent navy trainees requiring assistance with psychological problems from those not requiring such assistance.

---

[1] The opinions expressed are those of the writer and are not necessarily shared by the Department of the Navy.

For purposes of the study, situational variables were defined as those primarily in the present environment, while nonsituational variables were defined as those essentially independent of the present environment.

Procedure

The study was conducted at a training center for naval aviation technicians having approximately 5500 navy and marine enlisted men under instruction throughout the period of the study. This period was from July 1953 through June 1954.

As shown in Tables 1, 2, and 3, the seven situational variables consisted of the schools in which the students were enrolled and measures of success achieved by the students in the schools. The nine nonsituational variables consisted of aptitude test scores, overachievement-underachievement in a previous navy school, educational level, age, service membership (i.e., Navy or Marine Corps), sex membership, petty officer status versus nonpetty officer status, and recruits versus men returned from the fleet for training.

The study is actually two almost indentical studies, one for the period of the last half of 1953, and the other for the first half of 1954. It was decided in advance that the null hypothesis would be rejected only in those instances in which differences were significant at less than the .05 level in the case of both samples. It might be argued that the same end could be achieved by selecting a more exacting confidence level for the rejection of the null hypothesis, but this is not quite true. We know that conditions change to some extent both within the schools and with respect to the students assigned to the schools over a period of time. In the present study, we were interested in whether the changes that occurred over a period of several months would affect the significant differences observed in the initial sample, as well as in the stability of the initial findings from the standpoint of sampling error. Selecting a more rigorous confidence level would satisfy only the second of these two interests.

The psychological problems groups consisted of all men who were so classified by the medical staff at the naval dispensary to which they reported, either on their own initiative or were sent by the school which they were attending. Cases which were quite serious were transferred immediately to a naval hospital near the training center. Less serious cases were treated at the dispensary as outpatients. Thus, the psychological problems groups included students having a wide range of problems and behavioral disorders. The total number in the group for the last half of 1953 was 170, and for the first half of 1954, 162. This does not represent a decrease, since the population of the training center underwent a proportional decrease.

It was considered likely that certain schools were contributing a disproportionate share to the psychological problems groups. This proved to be true as shown in Table 1. This information, supplemented by the fact that number of students, aptitude test scores, and failure rates vary from school to school made it advisable to stratify the two samples of normal students with which the psychological problems groups were compared. The stratification resulted in the normal groups containing the same proportion of students from each school that the psychological problems group for the same time period contained. Having arrived at the proportion of the normal group that would be drawn from each school, the appropriate number was drawn by putting the IBM cards of all men attending the school during the period in service number order and drawing every nth card. If this stratification had not been accomplished it would have been difficult to state whether observed differences between the psychological problems groups and normal groups were due simply to school membership rather than to the variables per se. Two samples of 250 normal students were formed in this way, one from students attending the schools during the last half of 1953 and the other from students attending the schools during the first half of 1954. For purposes of the study, all students were considered to be normal if they had not been seen by the medical staff for psychological reasons. In the comparisons in which it was appropriate to compute means, the critical ratio statistic was used. In those instances in which the comparison of numbers of subjects was appropriate, chi square was computed.

Results and Discussion

The results of the study are shown in Tables 1, 2, and 3. Tables 2 and 3 contain all of the comparisons in which the chi-square

TABLE 1

SCHOOL MEMBERSHIP AND PSYCHOLOGICAL PROBLEMS

Aviation School	1953 Sample			1954 Sample		
	Observed	Expected[a]	Chi Square	Observed	Expected	Chi Square
Engines	13	32.8	14.8**	14	27.2	7.7**
Structures and Hydraulics	24	20.6	00.6	30	17.3	10.4**
Electronics	114	84.8	20.1**	100	81.0	8.9**
Training devices	11	10.2	00.1	10	9.2	0.1

[a] Expected frequencies are based upon the proportion of students in each school to total students at the training center.

** Significant at .01 level.

statistic and critical ratio statistic respectively were used, with the exception of school membership. In both of these tables the variables have been classified as being primarily situational or nonsituational.

School membership which is viewed as a situational variable is considered in Table 1. In the case of both the 1953 and 1954 samples, the aviation electronics school contributed a disproportionately large number of students to the psychological problems groups, and the aviation engines school contributed a disproportionately low number. It is interesting to note that the electronics course is more abstract and more difficult than the engines course. Furthermore, despite the assignment of students with higher aptitude to electronics, the course is failed by a considerably larger proportion of students than is the aviation engines course.

Table 2 indicates that both the 1953 and 1954 psychological problems groups contained a significantly larger proportion of academic failures than did the normal groups. From the data presented, however, one cannot be sure whether school failure tended to cause the higher incidence of psychological problems or whether a tendency toward psychological problems caused the observed higher failure rate. In addition to membership in certain schools and academic failure, a difference, significant at the .01 level, was found in the 1953 comparison on another variable classified as situational. This variable was school grades, as shown in Table 3. The same comparison for 1954 did not reach the standard set for statistical significance but closely approached it by

TABLE 2

COMPARISON OF PSYCHOLOGICAL PROBLEMS GROUP WITH NORMAL GROUP ON CERTAIN VARIABLES USING THE CHI SQUARE STATISTIC

Comparisons	1953 Groups					1954 Groups				
	Psychological Problems		Normal		Chi Square	Psychological Problems		Normal		Chi Square
	o	e	o	e		o	e	o	e	
Situational										
Academic failures	29	19	18	28	9.90**	24	17	20	27	4.79*
Other students	141	151	232	222		138	145	230	223	
Academic setbacks	84	79	112	117	.87	72	76	122	118	.75
Other students	86	91	138	133		90	86	128	132	
Nonsituational										
Petty Officers	13	11	15	17	.44	12	12	18	18	.01
Nonrated men	157	159	235	233		150	150	232	232	
Recruits	135	140	210	205	1.45	110	125	208	193	13.07**
Fleet returnees	35	30	40	45		52	37	42	57	
Navy men	110	123	193	180	7.86**	96	107	177	166	5.86*
Marines	60	47	57	70		66	55	73	84	
Male	164	166	245	243	.43	152	157	247	242	7.96**
Female	6	7	5	4		10	5	3	8	

* Significant at .05 level.

** Significant at .01 level.

reaching the .06 level of confidence. It should be noted that adequate data concerning school grades were not available for the students who failed to complete the schools. Hence the comparison on the basis of school grades includes only 67 of the 170 students in the 1953 psychological problems group and 42 of the 167 students in the 1954 psychological problems group.

The comparison involving the one remaining situational variable, academic setback, as shown in Table 2 did not yield a significant difference. The term "setback" refers to repeating a unit, usually two weeks in length, because of unsatisfactory progress in the unit.

Considering now the variables classified as nonsituational and on which significant differences were found in both the 1953 and 1954 samples, Table 2 indicates that the Marine Corps tended to contribute men to the psychological problems groups to a greater extent than did the Navy. Two additional variables classified as

TABLE 3

COMPARISON OF PSYCHOLOGICAL PROBLEMS GROUP WITH NORMAL GROUP ON CERTAIN VARIABLES USING THE CRITICAL RATIO STATISTIC

Comparisons	1953 Groups					1954 Groups				
	Psychological Problems		Normal		CR	Psychological Problems		Normal		CR
	Mean	*SD*	Mean	*SD*		Mean	*SD*	Mean	*SD*	
Situational										
School grades	73.15	5.00	75.25	5.64	2.89**	73.71	5.41	75.47	5.51	1.91
Nonsituational										
Marine aptitude battery, composite score	55.76	5.99	58.56	5.85	2.43*	55.14	5.88	56.53	6.96	1.25
Navy aptitude battery, composite score	57.16	5.18	58.65	4.71	2.46*	54.76	5.90	56.00	6.65	1.57
Over-underachievement index	98.97	5.23	100.14	4.91	2.00*	98.9	6.11	101.84	5.72	4.59**
Civilian education	11.21	1.21	11.82	1.27	4.77**	10.88	1.40	11.52	1.22	4.71**
Age	20.92	2.48	21.54	2.56	2.42*	20.94	3.67	21.35	2.90	1.18

* Significant at .05 level.

** Significant at .01 level.

nonsituational in Table 3 met the standard set for significance in both the 1953 and 1954 samples. The first of these was overachievement–underachievement in a previous service school, the psychological problems groups tending toward underachievement. The second variable was years of civilian education, the mean of the psychological problems groups being less than that of the normal groups.

The remaining nonsituational variables did not meet the condition set for significance for the study. These included aptitude scores, age, sex membership, petty officer versus nonpetty officer status and recruit versus fleet returnee. It should be noted in this connection that these results do not necessarily indicate that no differences exist on these variables. The conditions of the study were not sufficiently rigorous with respect to numbers and perhaps with respect to other considerations to justify the statement that any existing differences were given an adequate chance to be observed. It is believed, however, that one can have reasonable confidence in the stability of the differences that satisfied the standard set for the study for the *rejection* of the null hypothesis.

Theoretical Considerations

Empirical data frequently are better understood in the light of theory. Similarly, improvement of theory depends to a major extent upon its exposure to empirical data. One of the more highly developed theories of behavior is that of Hull (**3**). Brown and Farber (**1**) have added certain constructs to Hull's system with a view to a more adequate handling of emotional behavior. Further, they have illustrated how their constructs fit into the larger system by considering phenomena commonly ascribed to frustration, in terms of their proposed method of conceptualization. The present study clearly does not lend itself to the rigorous procedures laid down by Hull for checking a segment of a theory and revising it in accordance with the results of the experiment. Nevertheless, it is of interest to see to what extent Hull's system, as expanded by Brown and Farber, gives unity to the observed functional relationships and facilitates understanding of them. Reference to Brown and Farber's diagram summarizing the relationships between antecedent conditions, intervening variables, and consequent conditions (p. 492) will be helpful in the following discussion. This diagram has been reproduced by Lambert also (**4**, p. 66.).

One of Brown and Farber's key explanatory concepts is frustration. They use the term to connote a hypothetical state or condition of the individual, as opposed to a frustrating event or external barrier. The term is viewed as an intervening variable between other constructs or intervening variables which ultimately are anchored both at the antecedent end and consequent or response end by observable conditions. In the present study, the antecedent conditions may be thought of as including the 16 variables on which the normal group and the psychological problems group were compared. As noted previously, six of these antecedent conditions were found to be significantly related to the consequent condition, that is, behavior which resulted in the individual being classified in the psychological problems group.

It is assumed that successful completion of the aviation schools was an important intermediate goal or means to an end for virtually all students. The technical schools are a usual route to advancement and status in the various aviation occupations. Moreover, most of the students had asked to be sent to the school in which they were enrolled in preference to some other aviation school.

Adapting the theory to the present data the intervening variable, "drive," D_O, is envisioned as a function of the desire to complete the school, and D_C as a function of some other desire or need. More than one of these competing drives, D_C, may have to be taken into account in a given situation. This concept of drive is a departure from Brown and Farber's diagram on two counts. First, drive is described in terms of desired goals instead of deprivation or noxious stimulation. Second, in the Brown and Farber diagram, D without subscripts is used and defined as "relevant drive." Hull has defined D as "strength of dominant primary drive operative in the primary motivation to action after the formation of the habit involved." The use of D_O and D_C insures that the appropriate drives will be applied to the habits involved, that is, to habit strengths having the same subscripts as the drives.

The intervening variables ${}_SH_{R_O}$ and ${}_SH_{R_C}$ represent habit strength of an ongoing response tendency directed toward achieving the goal of school completion, and habit strength of an ongoing response tendency directed toward achieving some other goal, respectively. When ${}_SH_{R_O}$ and ${}_SH_{R_C}$ are multiplied by their respective drives two additional constructs emerge, ${}_SE_{R_O}$ and ${}_SE_{R_C}$, which represent initial excitatory strength of the response tendency directed toward school completion and inital excitatory strength of the response tendency directed toward a goal other than school completion. The intervening variable "competitive inhibitory tendency," I_R, is envisioned as a function of blocking and nonreinforcement of effort to achieve the goal of successful school completion. "Frustration," F, the key intervening variable, results from the competition or incompatability of ${}_SE_{R_O}$ with both I_R and ${}_SE_{R_C}$.

At this point it may be well to see how well the six variables on which significant differences were found fit into the theoretical structure described thus far. Membership in the electronics school is associated with high I_R values and membership in the engines schools with lower I_R values since the electronics school has a much higher failure rate than does the engine school. Similarly, after drawing a sample of normal students stratified to contain the same proportions from each school as were contained in the psychological problems group, we confirm our notion that school failure is significantly related to membership in the psychological problems group. This would result in a larger value of I_R in the individuals that ultimately made up the psychological problems group than in the normal group. These larger I_R values would more nearly equal the value of ${}_{S}E_{R_O}$ in the psychological problems group than in the normal group, a condition which, according to the theory, increases the value of F. The results concerning the antecedent conditions, educational level, and overachievement-underachievement in a previous navy school can be explained in a similar way. It is reasonable that the behaviors which underlie both of these variables are related to success or failure in the aviation schools. The explanation of the final variable on which a significant difference was found, membership in the Marine Corps versus membership in the Navy, is seen in the larger value of F in the case of marines due to competition between the incompatible response tendencies ${}_{S}E_{R_O}$ and ${}_{S}E_{R_C}$. The present study provides no data on the subject, but there is considerable informal evidence to indicate that marines join the corps with the goal of becoming a member of a conventional marine combat unit. Many are not volunteers for aviation, as is true in the case of navy men. This results in a higher D_C value for marines, which, when multiplied by its appropriate habit strength, ${}_{S}H_{R_C}$, results in a larger value of ${}_{S}E_{R_C}$ for marines than for navy men. Similarly, the conditions that result in a larger value of ${}_{S}E_{R_C}$ may also result in a somewhat smaller value of ${}_{S}E_{R_O}$ than that of navy men. For some individuals ${}_{S}E_{R_C}$ may exceed ${}_{S}E_{R_O}$. The behavior resulting from this condition under different circumstances will be considered later. In any event, the competing values of ${}_{S}E_{R_O}$ and ${}_{S}E_{R_C}$ would be expected to be more nearly equal in the case of the marine group than the navy group, and hence a larger value of F would be expected.

Brown and Farber hold that F has two major effects. First, it increases the general level of motivation. This effect is integrated into Hull's theory by considering F as an "irrelevant drive," for which Hull has presented formulas. In the present adaptation of the theory "effective drive strength," $\overline{D}_O$, results from the combination of the drive to complete the school, D_O, and the "irrelevant drive due to frustration," $\dot{D}$. When ${}_{S}H_{R_O}$ is multiplied by this new drive strength, $\overline{D}_O$, the construct, "final excitatory strength of the response tendency directed toward completion of the school," ${}_{S}E_{R_O}'$, emerges. $\dot{D}$ also combines with D_C, which, through the same process as that just described, results in ${}_{S}E_{R_C}'$, which is the "final excitatory strength of the response tendency directed toward a goal other than school completion."

Adapting the theory to the present data the intervening variable, "drive," D_O , is envisioned as a function of the desire to complete the school, and D_C as a function of some other desire or need. More than one of these competing drives, D_C , may have to be taken into account in a given situation. This concept of drive is a departure from Brown and Farber's diagram on two counts. First, drive is described in terms of desired goals instead of deprivation or noxious stimulation. Second, in the Brown and Farber diagram, D without subscripts is used and defined as "relevant drive." Hull has defined D as "strength of dominant primary drive operative in the primary motivation to action after the formation of the habit involved." The use of D_O and D_C insures that the appropriate drives will be applied to the habits involved, that is, to habit strengths having the same subscripts as the drives.

The intervening variables ${}_SH_{R_O}$ and ${}_SH_{R_C}$ represent habit strength of an ongoing response tendency directed toward achieving the goal of school completion, and habit strength of an ongoing response tendency directed toward achieving some other goal, respectively. When ${}_SH_{R_O}$ and ${}_SH_{R_C}$ are multiplied by their respective drives two additional constructs emerge, ${}_SE_{R_O}$ and ${}_SE_{R_C}$, which represent initial excitatory strength of the response tendency directed toward school completion and inital excitatory strength of the response tendency directed toward a goal other than school completion. The intervening variable "competitive inhibitory tendency," I_R , is envisioned as a function of blocking and nonreinforcement of effort to achieve the goal of successful school completion. "Frustration," F, the key intervening variable, results from the competition or incompatability of ${}_SE_{R_O}$ with both I_R and ${}_SE_{R_C}$.

At this point it may be well to see how well the six variables on which significant differences were found fit into the theoretical structure described thus far. Membership in the electronics school is associated with high I_R values and membership in the engines schools with lower I_R values since the electronics school has a much higher failure rate than does the engine school. Similarly, after drawing a sample of normal students stratified to contain the same proportions from each school as were contained in the psychological problems group, we confirm our notion that school failure is significantly related to membership in the psychological problems group. This would result in a larger value of I_R in the individuals that ultimately made up the psychological problems group than in the normal group. These larger I_R values would more nearly equal the value of ${}_sE_{R_O}$ in the psychological problems group than in the normal group, a condition which, according to the theory, increases the value of F. The results concerning the antecedent conditions, educational level, and overachievement-underachievement in a previous navy school can be explained in a similar way. It is reasonable that the behaviors which underlie both of these variables are related to success or failure in the aviation schools. The explanation of the final variable on which a significant difference was found, membership in the Marine Corps versus membership in the Navy, is seen in the larger value of F in the case of marines due to competition between the incompatible response tendencies ${}_sE_{R_O}$ and ${}_sE_{R_C}$. The present study provides no data on the subject, but there is considerable informal evidence to indicate that marines join the corps with the goal of becoming a member of a conventional marine combat unit. Many are not volunteers for aviation, as is true in the case of navy men. This results in a higher D_C value for marines, which, when multiplied by its appropriate habit strength, ${}_sH_{R_C}$, results in a larger value of ${}_sE_{R_C}$ for marines than for navy men. Similarly, the conditions that result in a larger value of ${}_sE_{R_C}$ may also result in a somewhat smaller value of ${}_sE_{R_O}$ than that of navy men. For some individuals ${}_sE_{R_C}$ may exceed ${}_sE_{R_O}$. The behavior resulting from this condition under different circumstances will be considered later. In any event, the competing values of ${}_sE_{R_O}$ and ${}_sE_{R_C}$ would be expected to be more nearly equal in the case of the marine group than the navy group, and hence a larger value of F would be expected.

Brown and Farber hold that F has two major effects. First, it increases the general level of motivation. This effect is integrated into Hull's theory by considering F as an "irrelevant drive," for which Hull has presented formulas. In the present adaptation of the theory "effective drive strength," $\overline{D}_O$, results from the combination of the drive to complete the school, D_O, and the "irrelevant drive due to frustration," $\dot{D}$. When ${}_sH_{R_O}$ is multiplied by this new drive strength, $\overline{D}_O$, the construct, "final excitatory strength of the response tendency directed toward completion of the school," ${}_sE_{R_O}'$, emerges. $\dot{D}$ also combines with D_C, which, through the same process as that just described, results in ${}_sE_{R_C}'$, which is the "final excitatory strength of the response tendency directed toward a goal other than school completion."

The second major effect of F is to produce unique internal stimuli, S_F. These stimuli may have innately associated response tendencies, ${}_{S_F}U_R$, or they may acquire new associative connections, ${}_{S_F}H_R$, through processes of learning. These response tendencies are largely maladaptive and may bear little rational relationship to the objective toward which the thwarted drive was directed. These response tendencies, when multiplied by S_F, result in the final construct in the system, ${}_{S_F}E_R$, "strength of excitatory tendency associated with frustration."

At this point we have three final response tendencies, ${}_{S}E_{R_O}'$, ${}_{S}E_{R_C}'$, and ${}_{S_F}E_R$. Assuming threshold strength, the response which will occur depends upon the relative strength of the three final response tendencies. If ${}_{S}E_{R_O}'$ has the greatest magnitude, behavior directed toward completion of the school will occur. This may take the form of additional study, seeking assistance from instructors or classmates outside of class, and/or other adaptive behavior related to successful school completion. If ${}_{S}E_{R_C}'$ is the strongest of the three final response tendencies, the other two will be inhibited, and behavior directed toward a goal other than school completion will occur. In the case of marines who joined the corps with a desire to be a member of a conventional marine combat unit, it may take the form of behaving in a manner calculated to obtain a transfer to such a unit. If ${}_{S_F}E_R$ is greater than the other final response tendencies, the response will tend to be maladaptive. In a military organization, perhaps to a greater degree than in a civilian setting, this type of behavior is likely to result in a referral for psychological problems. Unlike Dollard et al. (**2**), Brown and Farber do not specify precisely which response, e.g., aggression, withdrawal, or primitivization, will occur. The theory does include the expectations that the presence of frustration will often result in the intensification of response, that reactions followed by a reduction in frustration will be reinforced, and that responses associated with frustration in one situation will tend to appear in other thwarting situations.

The above discussion obviously is no more than exploratory in adapting the Hull, Brown and Farber theory to the type of data described. Despite the numerous limitations, including explanation after the fact and an absence of quantification, it is interesting to note the apparent ease with which the observed relationships can be handled by the theory. It raises the important question of whether the method employed in the present study could be adapted to the task of providing some of the empirical data required for detailed systematic development of the Brown and Farber theory of frustration. Perhaps this question could be answered by one or two studies in which the antecedent conditions were selected in accordance with the best information currently available, and measured as carefully as possible. The consequent conditions used in the study are already suited to the theory, since in its initial form it does not attempt to specify the precise reaction to frustration that will occur.

Summary

Two groups of navy and marine trainees having psychological problems were compared with two groups of normal trainees on the basis of 16 variables. Significant differences between both psychological problems groups and their corresponding normal groups were found on six of the comparisons. Students of aviation electronics tended to have psychological problems in greater numbers than did other students. The reverse was true in the case of men studying aviation engines. Academic failure was associated with membership in the psychological problems groups. The three comparisons just mentioned were thought of as being primarily situational in nature.

The remaining three comparisons on which significant differences were observed were classified as nonsituational variables. These were: overachievement-underachievement in a previous navy school, years of civilian education, and membership in the Navy versus the Marine Corps. Thus no consistent tendency for maladjustment to be more closely associated with situational variables or with nonsituational variables was found.

The findings were discussed in terms of Hull's theory, as adapted to emotional behavior by Brown and Farber.

References

1. Brown, J. S., & Farber, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, **48**, 465–493.
2. Dollard, J., Doob, L. W., Miller, N. E., Mowrer, O. H., & Sears, R. R. *Frustration and aggression*. New Haven: Yale Univer. Press, 1939.
3. Hull, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.

4. Lambert, W. W. Stimulus-response contiguity and reinforcement theory in social psychology. In G. Lindzey, (Ed.), *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954.
5. Rotter, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.

Received March 28, 1957.

COMPARATIVE EFFECTIVENESS OF TELEVISED AND LARGE CLASSES AND OF SMALL SECTIONS[1]

LAURENCE SIEGEL AND F. G. MACOMBER

Miami University, Ohio

College enrollments for the next decade are expected to rise sharply with an anticipated shortage of well-qualified college teachers in most subject areas. These conditions will place additional pressures upon public institutions to increase class size. The development of educational television, and particularly of relatively low-cost closed-circuit systems, raises the interesting possibility of teaching large groups of students while overcoming some of the visual difficulties inherent in large class instruction.

Miami University is engaged in a research program directed toward the comparison of the relative effectiveness of large group instructional procedures with smaller conventional-type classes of approximately 30 to 35 students. Two of the general approaches to large group instruction under investigation are designated LC (large classes with direct visual contact between student and instructor) and TV (closed-circuit television) classes. This paper summarizes the major results obtained during the Spring Semester, 1956. Three problem areas were investigated:

1. Achievement as a function of assignment to experimental or control sections;
2. Students' attitudes about the instructor and course as a function of section assignment;
3. Students' end-of-the-semester attitudes about having received TV or LC instruction rather than conventional instruction.

Procedure

Subjects

The first two problem areas under investigation required the comparison of results obtained from experimental and control sections of the eight undergraduate courses listed in Table 1.

---

[1] This research was conducted under a grant to Miami University from the Fund for the Advancement of Education.

Each experimental section was taught either as a large class (primarily lecture) or by means of closed-circuit television. The control sections were taught by conventional procedures. Both experimental and control sections were taught by the same instructor and no known biasing factor was known to be operative as a determinant of student registration for the experimental or control section. Nevertheless, the relatively small sizes of the control groups prohibited sole dependence upon random assignment of students to balance the effects of extraneous variables. Consequently, the experimental and control groups for each course were deliberately equated on the basis of scores on the Cooperative English Test (Form T), Cooperative Mathematics Achievement Test (Form P), and total score on the American Council on Education Examination (1948 edition), as well as grade-point-average for the previous semester.

The matching procedure required that some of the students actually enrolled in the experimental and control sections of each course be excluded from the specific experimental and control groups under investigation. All decisions about the constituency of the experimental and control groups were made prior to the collection of research data. Neither the instructors nor the students were informed regarding the actual composition of the groups providing data for the Experimental Study in Instructional Procedures. The total enrollments in each course and the numbers of subjects included in the study are summarized in Table 1.

A full description of the matching data is given elsewhere (**2**). It is sufficient here to indicate that none of the mean differences between the experimental and control groups in any course on any of the matching variables was statistically significant. Equating of the experimental and control groups on the critical variables of ACE and grade-point-average was further supported by the lack of statistical significance of the chi squares between pairs of score distributions on these two variables. Furthermore, both groups in each of the eight courses were demonstrated to be equated both with respect to proportional sex distributions and proportional distributions by class standing.

Teaching Procedures

The specific instructional procedures within each course were not standardized. Rather, instructors were encouraged to develop

TABLE 1

TOTAL ENROLLMENT AND NUMBERS OF EXPERIMENTAL SUBJECTS IN EACH COURSE

Course Title	Number of Sections			Number of Students Enrolled in Sections[a]			Number of Subjects from Each Section[b]		
	TV	LC	Control	TV	LC	Control	TV	LC	Control
Foundations of Human Behavior	1		2[c]	103		74	97		62
Principles of Human Physiology	1		1	96		31	90		30
Introductory Sociology	1		4[d]	141		144	131		129
Human Biology	1	1		183	178		183	177	
Business and Government		1	1		119	35		105	30
Composition and Literature		2	2[e]		100	50		37	37
Essentials of Modn. Geography		3	3[e]		213	93		88	85
Elementary Psychology		1	1		78	30		77	30

[a] Enrollments in multiple sections were divided approximately equally between each of the sections.

[b] These are the numbers of students from each section actually providing data for the Study.

[c] Multiple sections taught by the same instructor.

[d] Multiple sections taught by different instructors, each of whom rotated TV presentations.

[e] Multiple sections taught by different instructors, each of whom also taught an experimental section.

those procedures which seemed best suited to the instructional situation, and which were best calculated to achieve the objectives of the course. Thus there were considerable differences in the manner in which the four televised courses were conducted. The Foundations of Human Behavior class met for two weekly periods of 90 minutes each; Sociology and Human Biology were convened for three weekly periods of 50 minutes each. The students in the experimental Physiology section met for three 50-minute televised presentations and for a two-hour laboratory period each week. The nature of the televised presentations also differed in the four courses. In both Biology and Physiology, the general method was that of lecture supplemented by numerous demonstrations and training aids. Five different members of the department participated in the televised presentations of Sociology. These presentations were approximately 30 minutes in length with the balance of the 50-minute period devoted to a discussion in each receiving room under the leadership of a regular staff member. In Foundations of Human Behavior, the presentations took a variety of forms, but with approximately one-third of each class period given over to discussion under the leadership of graduate assistants who served as monitors in each of the receiving rooms.

Similarly, there was no standard form of procedure in the large classes not utilizing television as a medium of presentation. The LC sections of English followed essentially the same procedure as that designed for the control sections except that only half as many written themes were required, and, of course, individual student participation was considerably limited by the size of the experimental groups. In Geography, the students received two lectures a week with a third period devoted to small group discussion. The remaining two LC courses were given as three 50-minute lectures per week.

The control sections of all courses were conducted as discussion groups with lectures and demonstrations inserted by the instructor whenever he thought it desirable.

This diversity of approach was believed desirable even though it prohibited direct comparisons between courses. Forcing the instructors to adhere to a specific and inflexible type of presentation in both their experimental and control sections would have prevented them from capitalizing upon the advantages inherent in the specific classroom situation and would have emphasized some of the deficiencies.

Evaluative Instruments

Two classes of instruments provided data for this series of investigations: achievement tests and attitudinal measures. These two classes of instruments are described separately below.

"Achievement" was operationally defined for the present sequence of studies as performance on the objective portions of course examinations administered for the purpose of assigning final grades. It is recognized that such examinations typically measure only a single dimension of achievement, i.e., subject-matter knowledge. Thus the present data are not germane to such important course objectives as "synthesis," "problem solution" and "critical thinking." The possibility of leakage of information about the test was overcome by simultaneous administration of each examination to the experimental and control sections of each course with one exception. (Simultaneous administration in Biology was impossible because the experimental and control sections were taught during successive semesters.) Estimates of reliability of each test (Kuder-Richardson Formula 21) ranged from .55 to .93 with a median of .88. To the extent that the assumptions implicit in this computation were not met, the resultant values are probably underestimates.

Four Thurstone-type scales were included in the attitude battery administered near the end of the semester to students in experimental and control sections. These scales yield quantifiable information about (*a*) students' appraisal of the effectiveness of the instructor in presenting material and handling the class (the I scale); (*b*) students' appraisal of the course content (the C scale); (*c*) students' attitudes about having received televised rather than conventional instruction (the TV scale); and (*d*) students' attitudes about having received large class instruction rather than conventional instruction (the LC scale). Each of these scales provided a nine-point continuum with 5.0 as the neutral position. Student judges were requested to rate the items included in a preliminary pool, and items were selected for inclusion in the final forms of these instruments on the basis of low Q value and the median of the distribution of judgments.

Split-half reliabilities (corrected for length of test) based on responses of a sample of 300 respondents were .91 for the I scale and .92 for the C scale. The intercorrelation between these two attitude scales was .52. The corrected split-half reliability for the TV scale was .89 and for the LC scale was .92 in a sample of 100 students.

Results

Conclusions based on the results obtained from a single semester's instruction are subject to the influence of a variety of sampling errors. Only four TV and four LC courses were investigated. Conclusions based on these data might not apply to other courses or even to the same content taught by other instructors. Also,

TABLE 2

COMPARATIVE ACHIEVEMENT IN EXPERIMENTAL AND CONTROL SECTIONS

Course	TV Phase				Course	LC Phase			
	Section	Achievement		*t* ratio		Section	Achievement		*t* ratio
		M	*SD*				M	*SD*	
Foundations of Human Behavior	TV	123.76	16.03	0.25	Business and Government	LC	70.03	8.58	0.70
	Control	123.11	15.05			Control	71.28	7.65	
Introductory Sociology	TV	110.09	20.91	0.24	Composition and Literature	LC	28.94	5.59	1.33
	Control	109.53	16.90			Control	30.78	4.74	
Physiology	TV	208.92	31.33	0.66	Essentials of Geography	LC	79.83	10.59	0.83
	Control	204.59	26.44			Control	78.54	9.56	
Human Biology	TV	253.97	15.21	3.00*	Elementary Psychology	LC	109.45	15.93	1.77
	Control	248.66	15.76			Control	102.52	21.77	

* Significant at better than the .001 level.

results obtained from the present samples of students may not generalize to other samples enrolled in the same courses chiefly because of the relatively small size of some of the control groups.

Achievement

The distributions of achievement test scores in the experimental and control sections as well as appropriate *t* ratios are summarized in Table 2. The null hypothesis is not refuted for any course except Biology, i.e., with the exception of this one course, comparable achievement resulted from experimental and control presentations.

Interpretation of the statistically significant finding in the case of Biology is confounded by the fact that the experimental and control sections of this course were given during successive semesters rather than simultaneously. Although a definitive interpretation of the results obtained for Biology is impossible at the present time, it is apparent that neither TV nor LC presentations adversely affected achievement as defined by subject matter examinations. This finding for TV courses is in agreement with results reported earlier by the Pennsylvania State University (**1**).

The absence of differential achievement in experimental and control sections does not negate the possibility of differential achievement as a function of academic ability in interaction with

TABLE 3

COMPARATIVE ACHIEVEMENT AS A DUAL FUNCTION OF SECTION ASSIGNMENT AND LEVEL OF ABILITY

Course	Section	High Ability Subgroup				Low Ability Subgroup			
		Achievement Test			*t*-ratio	Achievement Test			*t* ratio
		M	*SD*	N^a		M	*SD*	N^a	
Foundations of Human Behavior	TV	126.28	29.74	43	0.34	115.81	16.06	43	1.37
	Control	128.38	13.92	26		120.96	14.77	29	
Introductory Sociology	TV	117.80	13.13	55	0.97	103.04	14.67	46	0.12
	Control	115.52	12.60	67		102.70	9.54	40	
Physiology	TV	221.70	29.34	37	0.41	197.87	29.20	39	0.25
	Control	217.73	19.27	11		195.76	27.28	17	
Human Biology	TV	258.03	13.81	95	2.20^c	249.56	18.84	86	2.61^c
	Control	253.14	16.50	98		241.45	17.12	68	
Business and Government	LC	70.86	8.06	49	0.81	69.02	6.93	51	0.31
	Control	72.86	8.14	14		69.80	8.89	15	
Composition and Literature	LC	32.43	2.79	14	0.06	26.73	5.57	22	2.09^b
	Control	32.33	5.29	15		29.73	3.95	22	
Essentials of Modn. Geography	LC	83.81	9.19	42	1.14	76.11	10.49	43	0.22
	Control	81.59	8.32	41		75.63	9.80	44	
Elementary Psychology	LC	113.02	13.98	44	1.03	104.53	17.15	32	1.75
	Control	108.61	17.69	18		92.55	24.06	11	

[a] Fluctuations from total N in the sample caused by unavailability of ACE scores or absence from final examination.

[b] $p < .05$. [c] $p < .02$.

section assignment. In order to investigate the possible existence of such interaction, the students within each experimental and control group were dichotomized with respect to ability on the basis of percentile conversion of total ACE store. The cutting point for this dichotomization was the fiftieth percentile. Comparative achievement of the "high ability" subgroups in the experimental and control sections and of the "low ability" subgroups in these sections is summarized in Table 3.

The data presented in Table 3 support the conclusion that level of ability does not interact with assignment to a TV section or its control counterpart as a dual determinant of achievement. A similar conclusion applies to all LC courses with the exception of Composition and Literature. The statistically significant difference for the "low ability" students in this one instance is the result of both a mean difference and a variance difference. It might have resulted from the unreliability of the final examination or from sampling errors. If replicated with a more reliable criterion and larger samples, it would indicate that this is one course wherein the less able students profit more from small class instruction than from LC presentations. The high ability students in this course, however, performed equally well under both classroom conditions.

Students' Attitudes About the Instructor and Course

It is desirable to camouflage the identity of specific courses and instructors for presentation of attitudinal data. These data are presented for courses designated by a letter rather than a title. The order in which the courses are hereafter listed differs from the order of listing in the previous discussions.

Group data summarizing the results obtained from administration of the I scale and the C scale in the experimental and control sections are summarized in Table 5. In courses taught by multiple instructors, the mean scores on both scales were obtained by pooling the data across instructors.

It is apparent from the data exhibited in Table 4 that TV instructors were rated at least as favorably by students in their television sections as by students in their control sections. Two of the four LC instructors, however, received significantly less favorable ratings in the experimental section than in the control section.

Referring again to Table 4, there is a pronounced tendency for students in the control sections to rate the course content more favorably than students in either the TV or LC sections, although the obtained difference is statistically significant in only three of the seven courses for which data were available. Apparently, instruction given simultaneously to large groups of students (either by TV or in large classes) is accomplished at some sacrifice of the students' favorability to the course even though it does not affect their examination performance. Although the data are not amenable to direct comparison of TV and LC instruction, they lead to the suspicion that this deficiency may be more marked in the case of LC instruction than it is for TV instruction.

Attitudes About TV and LC Instruction

Group responses to the TV and LC scales are summarized in Table 5. The fiducial limits extending the mean allow for an absolute interpretation of this value in terms of favorability or

TABLE 4

SUMMARY OF I-SCALE AND C-SCALE RATINGS IN EXPERIMENTAL AND CONTROL SECTIONS OF EACH COURSE

Course	Section	I Scale[a]			C Scale[b]		
		M	*SD*	*t*-ratio	M	*SD*	*t* ratio
A	TV	2.80	1.00	—	3.08	0.50	—
	Control[b]	—	—		—	—	
B	TV	4.56	1.74	1.35	4.50	1.24	2.27*
	Control	4.10	1.40		4.00	0.91	
C	TV	4.16	0.88	2.86**	4.39	1.26	0.94
	Control	4.56	1.24		4.23	1.26	
D	TV	2.94	1.02	0.11	3.43	0.88	0.94
	Control	3.14	1.00		3.28	0.85	
E	LC	4.64	1.16	2.14*	3.91	1.07	3.18**
	Control	4.00	1.48		3.21	0.86	
F	LC	3.58	1.06	2.67**	3.83	1.07	2.88**
	Control	3.10	1.16		3.37	0.89	
G	LC	3.42	1.08	0.77	3.97	1.10	1.26
	Control	3.22	0.98		3.68	0.83	
H	LC	3.60	0.80	0.00	3.93	0.95	1.90
	Control	3.62	0.40		3.50	0.95	

[a] Scores below 5.0 denote ratings as "more effective than average;" the lower the mean score, the more favorable the rating.

[b] Data not available.

* Significant between the .05 and .01 levels.

** Significant at better than the .01 level.

TABLE 5

SUMMARY OF TV AND LC SCALE RESPONSES

TV Scale				LC Scale			
Course	M	*SD*	Fiducial Limits*	Course	M	*SD*	Fiducial Limits*
A	4.91	0.83	±0.14	E	5.74	0.72	±0.16
B	6.19	0.81	±0.18	F	5.45	0.84	±0.19
C	5.42	0.99	±0.20	G	5.35	0.81	±0.26
D	4.76	0.94	±0.20	H	6.03	0.64	±0.13

* Fiducial limits of the mean calculated for the .05 level of confidence.

unfavorability to TV and LC instruction (taking 5.0 as a neutral position).

Two conclusions follow from these data:

1. It is impossible to generalize on students' attitudes about TV

instruction. Students in one course preferred TV instruction to conventional instruction (Course D); those in another course considered the two types of presentation to be of approximately equal effectiveness (Course A). However, students in two of the TV courses (B and C) felt that television was inferior to conventional instruction.

2. Students in all LC courses would have preferred assignment to a small section. The degree of this unfavorability, however, varied from slight (Courses F and G) to considerable (Course H).

The apparent variability in TV and LC attitudes from course to course is indicative of the fact that factors other than section assignment play a role in conditioning students' attitudes about the mode of instruction. The potential influence of two such factors (students' ratings of the instructor and academic ability) was investigated. Correlations between both I scale responses and ACE total score with responses to the TV and LC scales are exhibited in Table 6. Note that significant correlations involving the ACE are indicative of an inverse relationship, i.e., the higher the ACE score, the less favorable the TV or LC attitude.

It is apparent that if any ability subgroup is unfavorable to TV or LC instruction, it is more likely to be the high ability students than the low ability students. However, the inverse relationship between academic ability and attitudes toward the mode of presentation may be overcome by factors which are not yet entirely clear. One such mitigating factor is probably student perception about the instructor. There is a marked tendency for academic

TABLE 6

CORRELATIONS BETWEEN I-SCALE RESPONSES AND ACE TOTAL SCORE WITH ATTITUDE TOWARD EXPERIMENTAL TYPES OF INSTRUCTION

TV Phase			LC Phase		
Course	I Scale	ACE	Course	I Scale	ACE
A	.25[b]	.00	E	.32[b]	.06
B	.10	.32[b]	F	.13	.02
C	.17	.23[a]	G	.10	.38[a]
D	.36[b]	.14	H	.24[a]	.04

[a] r exceeds 1.96 times standard error of r of 0.00 ($p < .05$).

[b] r exceeds 2.58 times standard error of r of 0.00 ($p < .01$).

ability to correlate inversely with TV and LC attitudes in those courses where ratings of the instructor do *not* correlate with these attitudes.

Conclusions

1. Acquisition of subject-matter knowledge as measured by objectivity retest performance was not adversely affected by assignment to large (experimental) section (TV or LC) rather than to a small (control) section.

2. Academic ability did not interact with type of section assignment (experimental or control) as a complex determinant of achievement in any of the televised courses. A similar conclusion applies also to the LC courses under investigation with the possible exception of Composition and Literature.

3. Student ratings of the instructor (I scale) were not adversely affected by television. The only statistically significant difference between ratings by the TV and control sections occurred in one course wherein the instructor was rated more favorably by his TV section. Ratings of the instructor did tend, however, to be less favorable in LC sections than in control sections. The mean difference was statistically significant in two of the experimental courses.

4. Student ratings of the course (C scale) were less favorable in both TV and LC sections than in the control sections. This deficiency appears to be even more serious in LC sections than in TV sections (although data amenable to a direct comparison between these two types of presentation were not available).

5. Students' attitudes about TV instruction were not uniformly held across courses. In contrast, students in all LC courses were unfavorable to the mode of instruction when compared to small class instruction.

6. When academic ability was related at all to students' TV and LC attitudes, the direction of this relationship was inverse, i.e., high ability students were less favorable than low ability students. However, the fact that such a relationship was not present in all courses suggests the likelihood that other factors interact with academic ability in this regard. One such interacting factor is probably the students' attitudes about the effectiveness of the instructor.

References

1. Carpenter, C. R., & Greenhill, L. P. *An investigation of television for teaching university courses.* University Park: Pennsylvania State Univer., 1956.
2. Macomber, F. G., & Siegel, L. *Progress report of the experimental study in instructional procedures.* Oxford: Miami Univer. 1956.

Received April 25, 1957.
Early Publication.

BOOK REVIEW

D. B. KLEIN. *Mental Hygiene: A Survey of Personality Disorders and Mental Health.* New York: Henry Holt, 1956. Pp. 654.

This is the first revised edition of a book which appeared in 1954. It follows the same general plan and seeks to follow the committee which considered ways of improving the teaching of psychology, especially in emphasizing intellectual development rather than an applied interest. This book, therefore, is not merely a book on personal adjustment but deals more fundamentally with principles of mental hygiene.

Part One, Chapters 1 and 2, discusses the nature and scope of mental hygiene and gives a summary of its history and problems. Part Two, Chapters 3 and 4, deals with the nature and kinds of mental disorder. Part Three, Chapters 5 through 9, takes up the most important problem of the prevention of mental disorder. Part Four, Chapters 10 through 18, on the promotion of mental health, includes discussions of the home and the personality; developmental problems from infancy through adulthood; the dynamics of conscience; fixations; coping with reality; problems of repression, conflicts, frustration, anxiety, emancipation, and self-control; economic problems, security, morale, and the aged; two popular forms of idolatry—grades and competitive success—and the contribution of the right kind of religion to mental health.

"There are many ways to achieve personality integration, many ways to keep an even keel through the storms of life." There are many individual differences. There are other and perhaps better ways than the "hazardous trail of the go-getter." Recognition of the proper values in life may help us to avoid the "exaggerated devotion to dollar-chasing as the chief virtue of daily life." It may help us to a healthier life than that of "relentless competition and power-seeking and self-aggrandizement."

Although the book is not written primarily from the point of view of applied psychology, but rather from that of fundamental principles, the reader can find many practical applications in many parts of the book. Especially, he should be helped to a clearer understanding of psychology, particularly in its relation to mental health; and he can do so without being lost in technical language and unnecessary, complicated jargon, unverified assumptions, and reified elements, that are prevalent today and often pass under the name of psychology. It is generally clear and simple in presentation, and a glossary (pp. 627–642) adds distinctly to the usefulness of the book.

A few minor adverse criticisms seem relatively unimportant in comparison with the general excellence of the book. It is sufficiently comprehensive for a semester college course and can be understood by the beginning student in psychology. It will stand easily with the few really excellent books in its field. A good index (pp. 643–654) concludes the book.

A. S. Edwards

The University of Georgia

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 48 November, 1957 Number 7

CONSISTENCY OF SOCIOMETRIC STATUS POSITION IN MALE AND FEMALE HIGH SCHOOL STUDENTS[1]

RITA R. WERTHEIMER

Western Psychiatric Institute, University of Pittsburgh

The present study represents an attempt to investigate the consistency of male and female sociometric status position through the use of sociometric retests spaced at intervals of 12 months, 20 months, and 8 months. A separate evaluation of male sociometric status consistency and female status consistency for these same retest intervals is also included. The sociometric criterion situations employed are in the general area of leisure-time activities. Product-moment correlation coefficients were calculated to evaluate the consistency of the total sociometric choice frequencies obtained in the several testings.

Previous studies have shown that sociometric test scores have a very high degree of consistency, particularly when the retest interval is short. Jennings (**2**), as part of an extensive study of sociometric status in a training school for delinquent girls, reports a coefficient of .96 on a retest of one housing unit at the school after a four-day interval. Newstetter, Feldstein and Newcomb (**3**) found an average coefficient of .95 based on data given on four successive weeks with a five choice allowance on one criterion (tent mates in summer camps). Zeleny (**5**) reported, at the college

---

[1] This study was conducted while the author was a member of the psychology department of West Virginia University. Grateful acknowledgment is made to Quin F. Curtis, Chairman of the psychology department at West Virginia, Delmas Miller, Principal of University High School, and to the many teachers who cooperated in this project.

Bureau Edni. Researc

level using five choices and one criterion (membership in a discussion group), coefficients ranging from .93 to .95 on readministering a sociometric test on successive days.

These research findings as a whole suggest that a high correlation coefficient obtained by the method of immediate retest is probably indicative of a memory factor, as well as the slowness with which people change in their choice behavior.

With a longer retest interval, coefficients go down as would be expected on an a priori basis. Bronfenbrenner (**1**) in a study with the entire population of the University of Michigan laboratory experimental school (with a range of students from nursery school to the sixth grade) found that rank correlation coefficients between raw status scores obtained by each student in the fall and spring sociometric testing varied from .27 for the nursery school group to .59 for the sixth grade group. He concluded that in classroom situations marked shifts in sociometric status were comparatively rare and that, on the whole, children tended to retain the same general social position, with this tendency more pronounced in the older age groups. Jennings (**2**, p. 235) spaced a sociometric retesting at an interval of eight months' distance from the original testing and found with her sample of delinquent girls a reliability coefficient of .72 for leisure-time companion choices.

Present Study

Sample Characteristics

All of the tenth- and eleventh-grade boys and girls from one consolidated high school were used in the original sociometric testing in March 1953. Retesting of these adolescents occurred in March 1954 and again in November 1954. Since sociometric choices in this study were limited to members of home room groups, the home room assignment policy adhered to in this high school is pertinent. Typically, students were given home room assignments in the tenth grade, and they retained the same home room grouping and teacher until they graduated from the school or were otherwise separated from the institution. Because of this the home room grouping presents a fairly stable social situation which should be well structured by the spring of the tenth grade, and a significantly high relationship between sociometric status position in each of

TABLE 1

SAMPLE CHARACTERISTICS AT THE TIME OF EACH SOCIOMETRIC TESTING: NUMBER OF CASES, SEX, MEAN AGE, AND GRADE LEVELS

Characteristics	March 1953 Sociometric Test	March 1954 Sociometric Test	November 1954 Sociometric Test
Total Number	200	200	97
Males	103	103	50
Females	97	97	47
Mean Ages (in yrs.)	16.3	17.3	18.5
Number of Grade Levels			
Tenth	97	0	0
Eleventh	103	97	0
Twelfth	0	103	97

the testing periods should exist. Table 1 summarizes some relevant data about the subjects.

Sociometric Measuring Instrument

The same near-sociometric questionaire previously used by Wertheimer (**4**) on a comparable high school group was employed. This near-sociometric questionnaire had been constructed in terms of the age and experience of high school adolescents. It was evident that the criteria on which choices were to be made had to be meaningful to them in terms of their own interests and experiences. The questions used were confined to roughly comparable social situations. The attempt was also made to keep the questions on the same "level" of social rapport. For example, it was conceivable that an acceptable companion at a school party might not be the same one chosen for discussion of a serious personal problem Additionally, all of the questions were phrased positively so that the adolescents would not have their attention called to the less favored members of each group and, in this way, cause even more severe social maladjustment to these students.

The sociometric questionnaire that was used included the following instructions to the students:

> What we are going to ask you to do today is not a test of any kind, and will not become a part of your school record. There are no right or wrong answers to it. Anything that you include here will be confidential and will only be used as research material.

In response to the following questions, be sure and write down the three persons' first and last names. You may put down the same name in response to more than one question, if you would really choose the same person.

No additional specific instructions were given to the students. They were able to make companion choices from either sex in all three sociometric situations. The three criterion situations for which three choices were to be given were:

1. Suppose you have just won first prize on a quiz program. Among the prizes is an expense-free trip to Canada for yourself and three students *in your home room group*. Whom would you choose to go along with you?

a. ______________ b. ______________

c. ______________

2. Suppose you are planning a party at your home and have to limit the number of people you can ask to it. Which three members *of your home room class* would you invite?

a. ______________ b. ______________

c. ______________

3. Suppose you have a chance to spend the summer at a camp which has a section for girls and a section for boys. Which three members *of your home room group* would you prefer as fellow campers?

a. ______________ b. ______________

c. ______________

Administration of Questionnaire

Since all of the home room classes in the tenth and eleventh grades of the high school (in the March 1953 testing) were requested to fill in the questionnaire, it was necessary that its administration be closely geared to the school schedule. In consultation with the school authorities, it was decided that it could be most expeditiously administered by the teachers themselves. Accordingly, the experimenter met with them as a group, commented on the questionnaire procedure, possible sources of difficulty, etc. They were given the questionnaires in large envelopes that they were requested to seal in the presence of the students after administration and to return the material to the school office on the same day. With the exception of specific instructions to the students to limit their choices to their own home room groups, the teachers reported no difficulties in connection with the administration procedure.

The same procedure and sociometric questionnaire were used in the March 1954 and November 1954 administrations.

Sociometric Questionnaire Analysis

A roster of each home room class was obtained and a frequency tabulation made of the number of times each home room member was chosen by any other member as a companion for any or all of the criterion situations. Each person in the home room group obtained a total sociometric choice frequency based on three choices for each of the three leisure-time activity situations suggested in the questionnaire.

Treatment of the Data and Results

To evaluate the consistency of these total sociometric choice frequencies obtained by the designated students (both as a group and in terms of the male and female dichotomy) on subsequent sociometric testings, Pearson product-moment correlation coefficients were calculated between (*a*) March 1953 and March 1954 data for the total group and for the males and females separately; (*b*) March 1953 and November 1954 data for the entire group and for males and females separately; and (*c*) March 1954 and November 1954 data for the total group and for males and females separately.

The obtained correlation coefficients are presented in Table 2.

All of the nine coefficients are statistically significant beyond the .01 level and vary in size from .55 to .71. These data suggest that there is a high degree of consistency in sociometric status positon (as shown in expressed preferences for leisure-time com-

TABLE 2

PRODUCT MOMENT CORRELATIONS OF TOTAL SOCIOMETRIC FREQUENCIES OF MALE AND FEMALE HIGH SCHOOL STUDENTS IN THREE SOCIOMETRIC TESTINGS

	March 1953–March 1954 Comparisons			March 1953-November 1954 Comparisons			March 1954-November 1954 Comparisons		
Groups	*N*	*r*	*df*	*N*	*r*	*df*	*N*	*r*	*df*
Total	200	.55	198	97	.66	95	97	.68	95
Male	50	.56	48	50	.69	48	50	.67	48
Female	47	.62	45	47	.62	45	47	.72	45

Note.—All of these correlation coefficients are significant beyond the .01 level for the degrees of freedom indicated in each case.

panions) in adolescent high school students as a group and that this consistency is maintained by both males and females.

While this consistency in choice position may vary somewhat in different high schools (as, for example, in home room grouping systems that are more fluid than the system in the present study), it is likely that social status in leisure-time activities is well structured in the tenth grade and is probably accentuated but at least maintained until the last year in high school. These data further suggest that a pronounced group development had occurred by the tenth grade in the sample used in this study.

Summary

The present study investigated the consistency of male and female adolescent sociometric status position through the use of sociometric retests spaced at intervals of 12 months, 20 months, and 8 months. A separate evaluation of male sociometric status consistency and female sociometric status consistency for these same retest intervals was also included. The sociometric criterion situations employed were in the general area of leisure-time activities. Product-moment correlation coefficients were calculated to evaluate the consistency of the total sociometric choice frequencies obtained on the several testings, and all of these coefficients were statistically significant beyond the .01 level. The obtained results suggest that there is a high degree of consistency in sociometric status position (as shown in expressed preferences for leisure-time companions) in adolescent high school students as a group and that this consistency is maintained by both males and females.

References

1. Bronfenbrenner, U. The Measurement of sociometric status, structure, and development. *Sociometry Monogr.*, 1945, **6**, 3–80.
2. Jennings, H. H. *Leadership and isolation.* New York: Longmans, 1950.
3. Newstetter, W. I., Feldstein, M. J., & Newcomb, T. M. *Group adjustment.* Cleveland: Western Reserve Univer. Press, 1938.
4. Wertheimer, R. R. The relationship between sociometric status, sex, and socioeconomic level and Rorschach signs of adjustment with a group of adolescent subjects. Unpublished doctoral dissertation, Univer. of Pittsburgh, 1953.
5. Zeleny, L. D. Sociometry of Morale, *Sociol. Rev.* 1939, **6**, 799–808.

Received December 2, 1956.

AN EXPERIMENTAL STUDY OF THE EFFECT OF EXTERNAL DIRECTION DURING LEARNING ON TRANSFER AND RETENTION OF PRINCIPLES

JACK E. KITTELL

University of California, Berkeley

It is not unusual for two extremes to be associated with educational practice: (*a*) those teaching-learning situations containing such a minimum of structuring that they can be characterized as relatively independent discovery situations, and (*b*) those situations containing complete presentations of selected, specific principles and facts and characterized as rote memory situations. Various research findings have indicated that, perhaps, amounts of direction during learning that lie between those two extremes may be of value in promoting maximum learning, transfer, and retention when used in classroom situations. The results of this investigation support suggestions that some direction during learning aids discovery, transfer, and retention of principles.

The Problem

Experimental research has produced evidence that learner activity is of value in the discovery of principles. Classroom provisions for active pupil discovery have been found to be superior to methods emphasizing the specification of correct responses by the teacher in advance of discovery (**1, 10, 13, 14**). However, the extreme proposition that the more the freedom of learner activity and the less the amount of external direction, the better, as has been inferred in educational literature (**11,** p. 141; **15,** p. 80) is not necessarily a valid one. Various experiments have demonstrated that some degree of direction in discovery is superior to none at all in promoting transfer (**7, 8, 9**).

Three recent experiments somewhat similar in design and purpose to the one reported here have produced conflicting inferences. One experiment, using sixth-grade pupils as *S*s, produced results favoring independent discovery as superior to specification of correct response in advance of discovery activities (**12**).

Two other recent experiments used college graduates and college students as *S*s. The first of these two experiments found that of four different degrees of direction during discovery, ranging from a control group receiving no direction to a group supplied underlying principles but no correct responses, the latter was superior in promoting transfer (**4**). The other experiment compared the effects of supplying principles as direction and independent discovery on transfer and retention (**3**). No significant differences were found in the discovery of new principles. After an interval of one month the group supplied with principles during discovery retained significantly more learned principles than the control group.

The results of these experiments suggest that direction, in the form of underlying principles, supplied during discovery is beneficial in promoting transfer to similar problems and to retention after one month. It would seem important to conduct further experiments with elementary school pupils in situations somewhat comparable to actual teaching-learning situations to obtain evidence of the relative effects of Minimum, Intermediate, and Maximum amounts of direction during discovery on transfer to situations of varying degrees of differentness and on retention.

The hypotheses tested in the present experiment were that the amount of direction characterized as Intermediate (*a*) effects greater increases in learning and transfer to three situations of varying differentness, and (*b*) promotes greater retention through periods of two and four weeks.

Materials and Procedures

Subjects

The *S*s of this experiment, 132 sixth-grade pupils from public elementary schools in Pullman, Washington, were placed into three treatment groups through stratified-random selection. Stratification was based on high, medium, and low reading achievement classifications.

Final membership in the *Minimum* treatment group numbered 45, in the *Intermediate* treatment group, 44, and in the *Maximum* treatment group, 43.

Treatment group means and standard deviations of *S.R.A. Primary Mental Abilities* intelligence quotients and chronological ages are shown in Table 1.

TABLE 1

SRA INTELLIGENCE QUOTIENT AND CHRONOLOGICAL AGE MEANS AND STANDARD DEVIATIONS FOR THREE TREATMENT GROUPS

Statistic	Treatment Groups		
	Minimum	Intermediate	Maximum
Number of cases	45	44	43
IQ mean	107.0	105.98	106.30
Standard deviation	13.09	12.55	12.84
C. A. mean (months)	131.26	131.68	132.33
Standard deviation	4.55	4.48	5.06

TABLE 2

MEANS AND STANDARD DEVIATIONS OF NUMBER OF PRINCIPLES KNOWN ON PRETRAINING TEST FOR THREE TREATMENT GROUPS

Statistic	Treatment Groups		
	Minimum	Intermediate	Maximum
Mean	1.444	1.864	1.630
Standard deviation	1.500	1.540	1.100

The statistical difference among treatment-group IQ means were tested by referring the ratio of the variances among group means to the variance within groups to a table of the F distribution. The F of .063 was obtained indicating no significant differences among group means.

The distribution of chronological ages could not be accepted as normal and a nonparametric test, analysis of variance by ranks, was applied to test the significance of differences among the three treatment groups. The H value, 1.03, was found to be not significant. H, with the group numbers used in this experiment, approximates a chi-square distribution.

The mean number of principles known on the pretraining test and standard deviations for the three treatment groups are presented in Table 2.

The distribution of number of principles known on the pretraining test could not be accepted as normal, and again analysis of variance by ranks was applied to test the significance of differences among the three treatment groups. A resulting H of .0526 was found not to be significant.

On the basis of the above analyses it was assumed that the three treatment groups did not differ significantly in initial mental ability or pretraining knowledge of the principles used in the experiment.

Pretraining and Posttraining Tests

The pretraining and first posttraining tests for knowledge of principles consisted of 45 items grouped by threes for each of 15 organizational principles.

A second posttraining test, consisting of 45 items different from any used on other tests and grouped by threes for each of the same 15 principles, was used to measure transfer to new items based on previously experienced principles.

A third posttraining test, consisting of 45 different items grouped by threes for each of 15 different organizational principles, was used to measure posttraining ability to discover and use unlearned principles.

An example of the type of verbal problems used throughout this experiment follows: (*a*) Gone, (*b*) Start, (*c*) Go, (*d*) Stop, (*e*) Come.

The five words of each item were selected so that four were related by some principle and one did not belong. The principle was the reason why one word did not belong and it became the learner's task to discover the principle and apply it in order to discover the word not belonging. In the example the principle was that of paired opposites. "Start" and "Stop" are direct opposites as are "Come" and "Go" but "Gone" is of the wrong tense to be a direct opposite of any of the other four and so does not belong.

Twelve principles used in the pretraining, the training, and the posttraining tests were based on meaning relationships, one was based on sound relationships, one was based on familiar word combinations, and one was based on word arrangement and meaning relationships. If a wrong response was made for any item of a group of three, the *S* was given no credit for knowledge of the applicable principle.

The tests were distributed as booklets for group administration with all necessary instructions on the cover page. The directions stated that the items were grouped and spaced on the basis of underlying principles but gave no clue as to the nature of the principles. The *S*s were required to mark an X in the answer space corresponding to the word that did not belong with the other four in each item.

Seventy-six of the 180 items used in this experiment and 10 of the 30 principles had been previously used in a recent, somewhat similar experiment (**12**). All of the items used were composed of words selected from the first thousand of each of two well-known reading word lists (**5, 6**). Such selection provided reasonable assurance that the items were not tests of reading ability for the sixth grade pupils serving as *S*s.

All three tests used for measurement in this experiment were administered to 87 sixth-grade pupils in another community prior to the initiation of this experiment. An alternate form procedure was used to determine reliability and after correcting for length of one test, a reliability coefficient of .74 was obtained.

The third posttraining test composed of different items based on 15 different principles was compared with the first posttraining test for relative difficulty. The percentage of correct responses for each item was computed revealing a range from 6% to 72% with a mean of 35% for the first posttraining test and from 6% to 79% with a mean of 23% for the third posttraining test. The same procedure was used on the basis of principles with a resulting mean of 8% for the pretraining test and a mean of 2.3% for the third posttraining test. Five educators acting as judges supported the accepted assumption that the third posttraining test was more difficult than the first posttraining test.

Learning Materials and Treatments

Items used in the training tests were different but of the same type as those used for pre- and posttraining tests. The principles used for training were the same as those used in the pre-, first, and second posttraining tests.

Three different combinations of clues to the principles determining correct responses to the training items were incorporated into three sets of training materials, which when administered, each to a different group of *S*s, effected different treatments by providing three different amounts of direction during learning. The procedure of administration was consistent throughout the training period and followed practices previously tested by research. (**3, 4**).

The learning material for the group receiving Minimum direction consisted of items arranged as in the pre- and posttraining tests. Three items based on the same principle were grouped together and separated from other groups of items by spacing. The *S*s of the Minimum group were told that each group of three items was based on an underlying principle. No information was given this group as to the nature of the principles except that because of them one word in each item did not belong.

The group receiving Intermediate direction was provided with all the information supplied the Minimum group plus a verbal statement of the principle involved. The principles were printed immediately preceding the group of items to which they applied and were read orally twice with each administration of the items based on the principle. The stated principle for the item previously used as an example is: Form Two Pairs of Opposites.

The materials used by the Maximum treatment group included all the clues provided the Intermediate group plus oral statements of the three correct responses for each group of items. The oral statements of correct responses were made for one group of items at a time prior to the marking of responses by the *S*s. The oral statement made for the item previously used as an example was: "'Start' is the opposite of 'Stop' and 'Come' is the opposite of 'Go' so 'Gone' is the word that does not belong."

Knowledge of correctness or incorrectness of responses was provided all *S*s by the use of corrugated cardboard answer sheet backings, one surface layer of which was punctured to provide holes into which a pencil point dropped when the correct response for each item was marked. Such devices have been used successfully with groups in other experiments (**3**, **4**). Their use in the present experiment was accompanied by adequate proctoring and instructions which included a statement that *S*s should continue to mark answer spaces until the correct response was made, but that each incorrect response mark would lower the *S*'s total score and so no unnecessary marks should be made.

The training period was five weeks in length with nine items based on three principles being presented and repeated twice each week for a total presentation of 45 items based on 15 principles. To discourage *S*s' attempts to rely on memory of correct response patterns for the items presented a second and third time each week, the words in each item were varied in order for each presentation on the basis of five different correct-response patterns. The same patterns were used in punching holes in the answer sheet backings. The five different patterns were used in rotating order.

The pretraining test when administered the second time following training became the first posttraining test. The number of principles actually learned and transferred was the number of principles known on the first posttraining test not known on the pretraining test. Treatment-group means and variances of the number of principles learned and transferred to the first posttraining test are presented in Table 3.

Analysis of variance was used to test the significance of differences among the three treatment groups. The *F*, 4.73, was well within the .05 level of confidence. Accepting the hypothesis that significant differences existed among the three treatment groups, pairs of groups were compared through the use of a *t* test. Table 4 presents the differences between treatment-group means and the significance of *t* for each.

TABLE 3

TREATMENT-GROUP MEANS AND VARIANCES FOR NUMBERS OF PRINCIPLES LEARNED AND TRANSFERRED ON FIRST POSTTRAINING TEST

Statistic	Treatment Groups		
	Minimum	Intermediate	Maximum
Mean	2.555	3.660	3.744
Variance	4.290	3.570	4.440

TABLE 4

SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS OF TREATMENT GROUPS FOR PRINCIPLES LEARNED AND TRANSFERRED ON FIRST POSTTRAINING TEST

Group Pair	Difference Between Means	Standard Error of Difference	*t*	Level of Significance
Minimum and Intermediate	1.106	.422	2.620	P.02
Minimum and Maximum	1.189	.448	2.654	P.02
Intermediate and Maximum	.084	.429	.196	Not significant

Both the Intermediate and Maximum treatment groups were superior to the Minimum treatment group in learning and transferring principles to previously encountered situations. There was no significant difference between Intermediate and Maximum treatment groups in this situation.

Amount of Transfer to New Item Situations

The number of principles learned and transferred to new item situations was the number of principles known on the second posttraining test not known on the pretraining test. The mean number of principles learned and transferred and variances for each treatment group as measured by the second posttraining test are presented in Table 5.

An F of 18.87 was obtained through analysis of variance and found significant at the .001 level of confidence. The hypothesis that significant differences existed among the three treatment groups in the number of principles transferred to new item situations was accepted. The t test was applied to test significance of differences between pairs of treatment-group means. Table 6

TABLE 5

TREATMENT GROUP MEANS AND VARIANCES FOR NUMBERS OF PRINCIPLES LEARNED AND TRANSFERRED TO NEW ITEM SITUATIONS IN SECOND POSTTRAINING TEST

Statistic	Treatment Groups		
	Minimum	Intermediate	Maximum
Mean	1.930	4.590	2.907
Variance	3.520	4.160	5.190

TABLE 6

SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS OF TREATMENT GROUPS FOR PRINCIPLES LEARNED AND TRANSFERRED TO SECOND POSTTRAINING TEST

Group Pair	Difference Between Means	Standard Error of Difference	t	Level of Significance
Minimum and Intermediate	2.657	.415	6.40	P.001
Minimum and Maximum	.974	.444	2.19	P.05
Maximum and Intermediate	1.683	.458	3.68	P.001

presents differences between treatment-group means and the significance of t for each.

Inspection of Tables 5 and 6 reveals that the Intermediate treatment group was significantly superior to the other two treatment groups in the number of learned principles transferred to new item situations. The Maximum treatment group was superior to the Minimum treatment group in transferring principles to new item situations.

Posttraining Discovery of New Principles

The third posttraining test was composed of 45 items based on 15 principles. Neither items nor principles had been previously encountered by the *S*s during this experiment. The number of principles known on this test was determined by counting the number of principles for which *S* made three correct item responses. Treatment-group means and variances of the number of principles known on the third posttraining test are presented in Table 7.

The distribution of numbers of principles known on this test was considered questionable when the Chi-square Goodness of Fit test yielded a chi-square value of 15.03 found to be significant at the .02 level of confidence. Both a parametric technique, analysis of variance, and a nonparametric technique, analysis of variance by ranks, were applied to test the significance of differences among the three treatment groups as to the number of new principles discovered. An F of 38 and an H of 30.88 were obtained, both significant at the .001 level of confidence.

The t test and analysis of variance by ranks were both applied in testing for significance of differences between pairs of treatment groups. As both techniques produced values at the same levels of significance only the results of the t test are presented here.

TABLE 7

TREATMENT-GROUP MEANS AND VARIANCES FOR NUMBERS OF PRINCIPLES KNOWN ON THIRD POSTTRAINING TEST

Statistic	Treatment Groups Minimum	 Intermediate	 Maximum
Mean	1.733	4.590	2.674
Variance	4.20	4.67	4.93

TABLE 8

SIGNIFICANCES OF DIFFERENCES BETWEEN TREATMENT GROUP MEANS OF NUMBERS OF PRINCIPLES KNOWN ON THIRD POSTTRAINING TEST

Group Pair	Difference Between Means	Standard Error of Difference	t	Level of Significance
Minimum and Intermediate	2.857	.446	6.41	P.001
Minimum and Maximum	.941	.4525	2.08	P.05
Maximum and Intermediate	1.916	.468	4.09	P.001

Table 8 presents the differences between treatment-group means of the number of principles known on the third posttraining test and the significance of t for each.

The Intermediate treatment group was able to discover a significantly greater number of new principles than either of the other two treatment groups. The Maximum treatment group was able to discover a significantly greater number of new principles than the Minimum treatment group.

Number of Principles Retained

The pretraining test administered for a second time at the end of the training period as the first posttraining test was administered a third time two weeks following its second administration and two weeks later it was administered for a fourth time to measure the number of principles retained for two and four weeks. The number of principles retained was determined by the number known on the first posttraining test and also known on the two- and four-week retention tests.

The numbers of principles retained by each treatment group were compared establishing that the Intermediate group retained significantly more principles than the other two treatment groups for two and four weeks and the Maximum group significantly more than the Minimum group. However, more indicative of the relative effects of the amounts of direction on retention were the comparisons of the retention loss by each treatment group.

Analysis of variance by ranks for matched samples was used to test the significance of retention losses. Table 9 presents the treatment-group means of numbers of principles known on the

TABLE 9

SIGNIFICANCE OF DIFFERENCES FOR EACH TREATMENT GROUP BETWEEN THE NUMBER OF PRINCIPLES KNOWN ON THE FIRST POSTTRAINING TEST AND THE NUMBER REPEATED AFTER TWO WEEKS

Treatment Group	Mean Number of Principles: Originally Known	Mean Number of Principles: Repeated	χ_r^2	Level of Significance
Minimum	2.555	1.60	15.22	P .001
Intermediate	3.660	3.50	.82	P .40
Maximum	3.744	2.86	20.03	P .001

Note.—Analysis of variance by ranks, matched samples.

TABLE 10

SIGNIFICANCE OF DIFFERENCES BETWEEN THE NUMBER OF PRINCIPLES RETAINED TWO AND FOUR WEEKS FOR EACH TREATMENT GROUP

Treatment Group	Mean Number of Principles Retained: Two Weeks	Mean Number of Principles Retained: Four Weeks	χ_r^2	Level of Significance
Minimum	1.60	1.177	5.60	P .02
Intermediate	3.50	3.295	1.15	P .30
Maximum	2.86	2.300	11.12	P .001

Note.—Analysis of variance by ranks, matched samples.

first posttraining test, those retained after two weeks, χ_r^2 values, and the level of significance for each.

The Intermediate treatment group made no significant loss in retention of principles the first two weeks. Both the Maximum and Minimum treatment groups made significant losses in numbers of principles retained for two weeks.

Table 10 presents the treatment-group means of numbers of principles known on the two-week retention tests and repeated on the four-week retention test, the χ_r^2 values and the level of significance for each.

Both the Maximum and Minimum treatment groups continued to make significant losses in retention of principles while the Intermediate treatment group still had made no significant loss in the number of principles retained.

The significance of the losses of retention of principles by each

TABLE 11

SIGNIFICANCE OF DIFFERENCES FOR EACH TREATMENT GROUP BETWEEN NUMBERS OF PRINCIPLES KNOWN ON FIRST POSTTRAINING TEST AND THE NUMBERS REPEATED AFTER FOUR WEEKS

Treatment Group	Mean Number of Principles		χ^2_r	Level of Significance
	Originally Known	Repeated		
Minimum	2.555	1.177	20.00	P.001
Intermediate	3.660	3.295	2.75	P.10
Maximum	3.744	2.320	25.84	P.001

Note.—Analysis of variance by ranks, matched samples.

treatment group for the total four-week retention period is presented in Table 11.

The Intermediate treatment group made no loss in retention of principles during the four-week retention period that was significant at the .05 level of confidence. Both the other two treatment groups made significant losses in the retention of principles during both of the two-week retention periods and for the total of four weeks.

Discussion

In the present experiment, sixth-grade pupils benefited from clues which, in addition to organization of materials and information that such organization existed, provided relationships on which to base discovery.

Reasonable interpretation of all the evidence available indicates that *S*s benefit from help given them in their search for bases determining correct responses. Specification of answers in advance, however, seems to encourage reliance on rote memory rather than discovering and applying on the basis of underlying relationships. Evidence provided by the present investigation seems to support such indications.

The present investigation differed from previous experimentation in that the relative effects of Minimum, Intermediate, and Maximum amounts of direction on transfer to three different situations and on retention were investigated within a single design. The Intermediate amount of direction, which provided as clues to discovery statements of underlying principles, proved to be superior to a Minimum amount of direction in promoting learning and transfer to all three types of situations. The Intermediate amount of direction was superior to Maximum amount of direction in promoting transfer to all but one type of situation. Two and four weeks after training, the group receiving Intermediate direction retained significantly more principles known in that situation than did the groups receiving Maximum or Minimum amounts of direction. The evidence provided by this experiment indicates that both the Minimum and Maximum amounts of directions during learning became less and less effective as the situations to which transfer was made became increasingly different and as the elapsed time between training and measurement of transfer became greater. The Intermediate amount of direction increased in effectiveness between the first and second transfer situations and maintained the new level of effectiveness on the third transfer situation.

The scope of this experiment has acknowledged limitations. Should further research confirm the following implications, the teacher's role in providing direction in learning-teaching situations would be clarified.

Evidence provided by the present experiment suggests that:

1. In addition to organizing the materials used in learning, teachers should aid pupil discovery by suggesting meaningful relationships on which learners may base discovery and by providing practice with those relationships.

2. Providing statements of underlying relationships without specifying answers fosters learning, retention, and transfer to different situations.

Summary

The purpose of this experiment was to determine the relative effects of three amounts of direction to learners in their discovery of established principles on transfer to differing situations and on retention of learned principles.

Each of three groups of sixth-grade pupils was supplied a different amount of direction during discovery of principles determining solution of multiple-choice verbal items. The group receiving an Intermediate amount of direction consisting of organized materials, information of such organization, and statements of underlying

Edn!. Rese

relationships, learned and transferred as many or more principles to three differing situations than the groups receiving less or more direction. After periods of two and four weeks subsequent to training, the group receiving an Intermediate amount of direction retained a greater proportion of learned principles than the other two groups.

Evidence from this experiment in conjunction with that of similar experiments indicates that furnishing learners with information in the form of underlying principles promotes transfer and retention of learned principles and may provide the background enabling future discovery of new principles.

References

1. Anderson, G. L. Quantitative thinking as developed under connectionist and field theories of learning. In Esther G. Swenson, G. L. Anderson, & C. Stacey (Eds.), *Learning theory in school situations*. Minneapolis: Univer. of Minnesota Press, 1949. Pp. 40–73.
2. Brownell, W. A., & Moser, H. E. *Meaningful versus mechanical learning: a study in grade III subtraction*. Durham, N. C.: Duke Univer. Press, 1931.
3. Craig, R. C. Directed versus independent discovery of established relations., *J. educ. Psychol.*, 1956, **47**, 223–234.
4. Craig, R. C. *The transfer value of guided learning*. New York: Bureau of Publ., Teachers Coll., Columbia Univer., 1953.
5. Dolch, E. W. *Problems in reading*. Champaign, Ill.: Garrard Press, 1948.
6. Gates, A. I. *Reading vocabulary for primary grades*. (rev. ed.) Bureau of Publ., Teachers Coll., Columbia Univer., 1935.
7. Hendrickson, G., & Schroeder, W. H. Transfer of training in learning to hit a submerged target. *J. educ. Psychol.*, 1941, **32**, 205–213.
8. Judd, C. H. The relation of special training to general intelligence. *Educ. Rev.* **36**, 1908, 28–42.
9. Katona, G. *Organizing and memorizing*. New York: Columbia Univer. Press, 1940.
10. McConnell, T. R. Discovery versus authoritative identification. In T. R. McConnell et al. (Eds.), *Studies in the psychology of learning II*. Iowa City: Univer. of Iowa, 1934.
11. Spitzer, H. F. Learning and teaching arithmetic. In N. B. Henry (Ed.), *Yearb. nat. Soc. Stud. Educ.*, **50**, Part II. *The Teaching of Arithmetic*. Chicago: Univer. of Chicago Press, 1951.
12. Stacey, C. L. The law of effect in the retained situation with meaningful material. Unpublished doctoral dissertation, Univer. of Minnesota, 1945.

13. Swenson, Esther J. Organization and generalization as factors in learning, transfer, and retroactive inhibition. In E. J. Swenson, G. L. Anderson, & C. L. Stacey (Eds.), *Learning theory in school situations.* Minneapolis: Univer. of Minnesota Press, 1949. Pp. 9–39.
14. Thiele, C. L. *The contribution of generlization to the learning of the addition facts.* New York: Bureau of Publ., Teachers Coll., Columbia Univer., 1938.
15. Thiele, C. L. Arithmetic in the middle grades. In N. B. Henry (Ed.), *Yearb. nat. Soc. Stud. Educ.*, **50**, Part II. *The Teaching of Arithmetic,* Chicago: Univer. of Chicago Press, 1951. Pp. 76–102.

Received January 11, 1957.

THE PERSONALITY OF JUNIOR HIGH AND HIGH SCHOOL STUDENTS ATTENDING A RESIDENTIAL SCHOOL FOR THE BLIND

HERBERT M. GREENBERG, LOUISE ALLISON,

MILDRED FEWELL, AND CHARLES RICH

Texas Technological College, Lubbock

The method by which the blind can best and most profitably be educated has been a subject for controversy for many years. The impact on the personality of the blind child of the residential school which educates better than three-quarters of all blind children is widely debated in the field. Some work has been done regarding the intellectual effects (**5**), which in general has shown satisfactory intellectual growth for blind students. Nothing, however, has been done in the way of focusing on the school for the blind as an effecter of personality. It is the purpose of this study, through use of a sample from a representative school for the blind, to compare the norms of this group on several personality traits with the general norms of their sighted age mates. In other words, it is the purpose of this study to investigate a hypothesis that students attending a residential school for the blind will be: more neurotic, less self-sufficient, more submissive, less self-confident, about as sociable, and more authoritarian than the norm for the sighted group for these traits.

Procedure

Subjects

One hundred and three students, constituting practically all those attending grades six through twelve in a residential school for the blind, were tested. This school has two categories of students. By far the largest number live at the school for the approximate 9½-month academic year and return home for the summer. A few whose families live in the town in which the school is located return home daily or on weekends. Living at home is greatly discouraged by the school administration.

The students represent a cross section of backgrounds and a reasonably accurate cross section of the socioeconomic groupings of the state. There is a wide range of academic ability, with the norm probably being very similar to that of the sighted population. Those few graduates who attend college are generally quite successful, though it should be noted that this number is quite small.

Test Battery

The Bernreuter Personality Inventory, the California F Scale, and a personal data form were administered to the students by administrators working simultaneously with four different groups.

The Bernreuter Personality Inventory has been standardized on thousands of high school and college students and adults. Percentile norms have been established for each of these groups, and for males and females in each group. The traits measured by this instrument are: neurotic tendencies, (B1N); self sufficiency, (B2S); introversion-extroversion, (B3I); dominance submission, (B4D); self confidence, (F1C); social ability, (F2S). The B3I was not employed in the present study, due to its very high correlation with B1N and F1C. The Bernreuter consists of 125 items to which the respondent answers with a yes, no, or question mark.

The California F Scale was constructed by Adorno and others, and has been standardized on several thousand students and adults. It consists of 29 statements to which the *S* responds by various shadings of agreement or disagreement (strongly agree +3, agree +2, mildly agree +1, strongly disagree −3, disagree −2, mildly disagree −1). An example of the type of items employed is "familiarity breeds contempt."

A personal data form has been employed by the senior author in his doctoral dissertation (**4**). It was designed to take data pertaining to the background of the subject so that the establishment of comparison groups would be possible. It contains such items as age, sex, approximate yearly income of family, religious persuasion, length of time at school, satisfaction with school, grades, etc. It is not in any way to be deemed a psychological test of a standardized instrument.

Methods of Collecting Data

In order to eliminate so far as possible the administration as a variable, a standardized administration was used by all four.

A script was prepared, which was followed exactly by each, and thorough rules as to the type of questions that could be answered were made and followed. Only definitions of words could be given to the *S*s and no interpretation of test items was permitted. Those *S*s who were able responded on typewriters to orally dictated questions while others employed pencil or Braille.

The respondents were told not to put their names anywhere on the paper and were assured that their responses could not be traced in any way to them. It was felt that with this guarantee of anonymity more candor might be hoped for.

Methods of Treating Data

Means and standard deviations were found for each of the three subgroups: Subgroup A—6th, 7th, and 8th grades; Subgroup B 9th and 10th grades; and Subgroup C 11th and 12th grades. These groups were further broken down into male and female for the Bernreuter B scales due to the differential sex standardization of these scales on the inventory. The t ratios were then taken between means of each subgroup for each of the traits tested. Results for each question of the personal data form were tabulated for each of the subgroups.

Results

Table 1 gives the mean percentile for the several subgroups for each of the traits measured. From this table it can be seen that the group exhibited high neurotic tendencies. The lowest subgroup had a mean score at the 66th percentile, while the remainder were at the 80th and above. It can also be seen that the group is quite dependent. One subgroup has a mean as low as the 15th percentile, while the most self-sufficient subgroup had its mean at the 34th percentile.

It can further be seen that this trend is carried out in the dominance-submissive scale, the groups lying on the submissive side. This was the only trait besides the authoritarian in which significant *t* ratios were found between any of the subgroups. In this case it was between the youngest and the oldest males, *t* ratios being 3.15, or above the 1% level of significance.

From the high neurotic trend exhibited by the group, it might be assumed that there is a lack of self-confidence. This, in fact, is clearly indicated by Table 1 which shows that no subgroup

TABLE 1

RESULTS FOR THE SEVERAL TESTED TRAITS BERNREUTER SCALES EXPRESSED IN PERCENTILE FIGURES[a]

Group	B1-N	B2-S	B4-D	F1-C	F2-S	F scale
Male A	82	15	15	81	29	152
Female A	66	22	41	81	29	152
Male B	83	19	23	86	25	167
Female B	83	17	37	86	25	167
Male C	81	21	46	80	45	151
Female C	80	34	54	80	45	151

[a] The first five traits on the above table are scales of the Bernreuter Personality Inventory taken from the manual of that inventory. The sixth trait refers to the California F scale.

B1-N—Neurotic Tendencies; B2-S—Self Sufficiency; B4-D—Dominance-Submission; F1-C—Self Confidence; F2-S—Social Ability; F scale—Authoritarianism.

has a mean below the 80th percentile, thus demonstrating a severe lack of self-confidence.

It should be noted that the trait of sociability is the only one in which the group lies on the healthy side of the norm. The most unsocial subgroup having a mean percentile of 45, however, authoritarianism is carried out in the pattern set by most traits, the means ranging from 151 to 167 on a scale ranging from 29 to 203, the high numbers representing high authoritarianism. It should be further noted that the 9th and 10th graders show a significantly more authoritarian score than the 6th, 7th, and 8th graders, the *t* ratios being 4.71 or again well above the 1% level of significance.

The most noticeable result obtained by the personal data form was the apparent widespread dissatisfaction with the school. In Subgroup A, 27 would choose this school again if having to choose again, while 17 would not; in Subgroup B, 13 would choose again while 17 would not; and in Subgroup C, 16 would choose again and 17 would not. Also, many who answered affirmatively did so because they felt that there was no alternative but expressed great hostility toward the school. The subgroups answered the questions quite consistently with one another. There were no differences in the responses to the other questions on the personal data form as great as the difference indicated in the above para graph except, of course, for the obvious differences of age and length of attendance at the school.

Conclusions

It is concluded that there is poor mental health at this residential school for the blind, as in most of the traits tested the mean percentile is far over on the unhealthy side of the scale, thus confirming the hypothesis. The researchers might pose two alternative causal possibilities. The personality difficulties might be due to the sensory loss itself, or to some peculiarity of the particular school. It should be noted in this connection that, in previous research (**4**), scores still decisively unhealthy but more closely approaching the norm were found in other residential schools, while scores clearly on the healthy side of the scale were found for the blind attending public schools. Thus some weight might be lent to the second alternative.

References

1. Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. Bernreuter, R. G. *Manual of personality inventory*. Stanford: Stanford Univer. Press, 1935.
3. Cutsforth, T. D. *The blind in school and society*. New York: D. Appleton, 1933.
4. Greenberg, H. *Some effects of segregated education on various aspects of the personality of those members of disadvantaged groups experiencing this form of education*. Unpublished doctoral dissertation.
5. Komisar, D., & MacDonnell, M. Gains in IQ for students attending a school for the blind, *Except. Child.*, 1955, **21**, 127–129.

Received March 28, 1957.

A PROCEDURE FOR MANAGING TEACHING-LEARNING SITUATIONS IN EDUCATIONAL PSYCHOLOGY

Ray H. Simpson

University of Illinois

Can the teacher of educational psychology seriously attempt to practice what is preached? Can the course in educational psychology, or any psychology course for that matter, give daily evidence to students that knowing psychological theory and principles really makes a desirable difference in the way teaching-learning situations are handled? In his own classes dare the teacher of psychology try to use what he knows in guiding the work of his own students? Most teachers of psychology have probably given some attention to questions such as these. The folowing description represents one attempt developed over a long period of years to answer these questions in the affirmative. It is not given as a model but rather as an approach which may be suggestive to others who are also concerned about possible ways of translating psychological theories and principles into practice in realistic teaching-learning situations.

Course Objectives

The general purpose of the course in educational psychology, as seen by the writer, is to prepare each student to do effective and satisfying professional work through the identification and application of psychological principles. Learning the attitudes and skills that will be needed for professional growth while teaching is emphasized; memorizing printed materials for test purposes is discouraged except as this can be shown to contribute to professional growth. Since the students in the current educational psychology course expect to teach, the attempt will be made to: (*a*) focus attention on the improvement of teaching and learning, (*b*) provide for cooperative consideration by the teacher and the learners in the class of the various processes involved in improving the teaching-learning situation, and (*c*) encourage the understanding and frequent use of varied methods of attacking professional problems in teaching-learning situations.

Course Activities

The writer's purposes in the educational psychology course are revealed in the outline of the major course activities, which are discussed in this paper: (*a*) identifying some key problems in teaching-learning situations, (*b*) selecting problems for initial study, (*c*) working toward problem solution, (*d*) keeping a student learning record, (*e*) self-evaluation, (*f*) finding and using resources, and (*g*) developing effective interpersonal relations.

With the guidance of the instructor, each student considers a great number of teaching-learning problems or situations. For example, consideration is given to such questions as these: On what basis should we group in this class? Should the total class group be used for all purposes? What use, if any, should be made of small groups? For what purposes might small groups be formed? Should out-of-class study groups be encouraged? What criteria should be used in marking in order to promote desired learning? These and similar questions relating to situations both in and out of the class vary with respect to implied educational objectives and methods. Some goals and activities are those of individuals, some of small groups, and some are those common to the whole class. In this class much attention is given to helping students develop sound goals and purposes which were not initially recognized.

Identifying Problems

In order to provide a sound basis for later individual study, each learner, during the first few days of the course, is encouraged to identify and list a number of professional problems that seem to be of concern to him. Those who have had teaching experience are encouraged to draw on this experience. Students without teaching experience may object that they do not know any professional problems. Actually, most students will have had a minimum of 13 years in teaching-learning situations. Through instructor guidance and some supervised observations of elementary and secondary school classrooms, the writer has found that undergraduate students without teaching experience show considerable ability in identifying teaching problems. However, direct experiences are not relied on exclusively.

Students are encouraged to use many types of resources in widening their perception of the important issues now facing educators and of the problems they are likely to meet as teachers. Recommended resources for problem identification include: direct observation of classrooms taught by other teachers, educational psychology texts, other printed materials—particularly in educational and psychological magazines—other students and their difficulties, and other teachers and their problems. The following are problems illustrative of those listed by students:

How can I more adequately adapt my instruction to the individual differences in classes in ninth-grade general science? (Origin: own experience as a student in general science class which was not challenging to me.)

What techniques can a young and inexperienced administrator use to encourage older teachers to modify their practices? (Origin: talking with a recent college graduate who was somewhat discouraged.)

What are some of the guiding principles and procedures that can be used in working with slow learners in the classroom? (Origin: table of contents of an educational psychology text.)

What is good discipline and how can I promote it in my classes? (Origin: Experience as a student in classes. Also difficulties of a teacher in classes observed while I was a college sophomore.)

After the learners have compiled their lists, they are encouraged to get together in small groups to select the problems they consider most important. Sometimes this is done during class time if it is difficult for groups to meet outside of class. The goals here are to help each learner expand the scope of his perceived problems, to get the reaction of others, and to appreciate the kinds of problems which others see as important. It is believed that activities will help the students in their approach to the professional problems they will later encounter. Conventionally, of course, this is done, if it is done at all, almost entirely by the teacher.

Selecting Problems for Initial Study

After each student has identified a number of professional problems which he might consider studying (usually between 15 and 35), all are encouraged to help set up criteria which will aid in deciding which most deserve attention and study. The instructor usually suggests one or two questionable criteria and one or two better ones in order to get learners started, e.g.:

Questionable: What is easiest to teach or learn without regard to importance? What problem was studied first last year?

Better: What are the present and prospective needs of individual learners in the class? (It is hoped that the students will eventually encourage their pupils to consider similar criteria.) What kinds of needed psychological training or practice is each learner not likely to get in other phases of his educational preparation?

After hearing an example or two of each type, each student tries to list both desirable and questionable or overemphasized criteria. He is encouraged to give his own reactions, not to try to anticipate the instructor's attitudes. Subsequently, the ideas are pooled, and a list is put on the blackboard which involves the combined judgments of the class members. Students are encouraged to consider the significance of particular criteria in deciding what problems it is best to study.[1] The criteria which are judged to be the most valid are then used in deciding which problems will be studied intensively. The process just described helps each student develop basic purposes for his study and tends to deemphasize excessive emphasis on working merely for a mark.

After the students have idendified a number of problems on the basis of the criteria previously listed, each, with instructor help, is encouraged to select two, three, or four problems and break each of them down into subproblems. An illustration of a problem with its subproblems that one student developed initially is the following:

How can I meet with a greater degree of adequacy the problem of individual differences in my classes in ninth-grade general science?

1. What are my specific purposes in teaching the course?
2. What are probable learner goals in taking the course and what are possible ways of determining these?
3. What plans should I consider making before the classes meet?
4. What are ways of determining a learner's science needs?
5. What are possible organizational setups, logical and/or psychological, which might be used in developing the course?
6. What are psychological strengths and weaknesses or limitations of various class procedures which other science teachers have proposed?
7. What procedures tend to develop critical thinking and the application of the scientific method to science problems in individuals of differing ability?
8. What are possible ways of evaluating course work and helping learners evaluate their work?

---

[1] For additional development of issues related to problem identification see: Ray H. Simpson, *Improving Teaching-Learning Processes*. New York: Longmans, Green, 1953, pp. 41–96.

9. What are ways of making group work more effective in promoting learning?

10. What kinds of resources will we be likely to need and how should these be obtained?

After the student has broken down two, three, or four problems into specific questions or subproblems—each problem with its subproblems constituting a possible working outline—the teacher and other learners in the group offer suggestions for improving them. From this group of two, three, or four tentative outlines, the student is in a position to decide which problem with its subproblems it would probably be best for him to study. After his decision has been made, the student may add, delete, restate, or reorder the subproblems. This refined statement of a major problem and its subproblems constitutes his basic organization or psychological frame of reference for early study. Several problems may be attacked by a student during the course of a semester.

Working Toward Problem Solution

After the learner has organized his subproblems into a tentative outline, he is in a position to start looking for solutions and to react critically to ideas which are gleaned from various resources. It is emphasized by the instructor that for any one educational problem there is no single "best answer" since the teacher and the learners differ significantly in their needs or abilities in different situations. For example, different approaches to discipline are usually called for in a rural neighborhood and in an urban slum district.

An attempt also is made to help the learner see that there are usualy arguments for and against each proposed solution, and he should become familiar with these arguments if he is to discuss the issues involved intelligently. Important in the understanding of the arguments is a consideration of the psychological principles or theories upon which each is based. A study of these principles and theories helps the learner build up a theoretical framework upon which he may make more valid judgments in other realistic problem situations.

Keeping a Student Learning Record

In the class in which a specific attempt is made to encourage the development of more effective learning processes, it is important that a rather complete record be kept by the student of his learning activities and ideas. Such a comprehensive record is useful from several standpoints. First, it enables the student to clarify his plans and ideas and the information acquired by writing them down in his own words. Second, it gives the teacher and other students an opportunity to examine and offer suggestions about the progress of his work in the course. Third, it provides an objective basis upon which, with the guidance of the teacher, he can evaluate the work that he has done to date and make plans for future activities.

The students are usually encouraged to keep a record of the following sort. Each should have an orientation section which includes a list of the problems which he has identified as important from his standpoint. This section should also include a breakdown of several problems which he has considered for initial study. In the orientation section of the learner's record there is also a list of the comments, criticisms, or suggestions which have been made by the instructor or by other students. The orientation section is usually followed by an outline of the main problem which he is attacking together with the subproblems under it.

Each learner keeps a resource sheet on which he lists, usually by number, all resources which he has found useful in connection with the course. After completing the resource sheet, the learner prepares a number of pages at the top of each of which he writes down one subproblem. As he finds possible solutions of his subproblems, he lists these on the appropriate pages. The origin of each possible solution is indicated by a number which refers to the location of the item in the bibliography or on the resource sheet.

Under each subproblem, after each group of ideas from a specific source, the learner writes his own critical comments and reactions to the ideas he has obtained from various sources. The primary purpose of encouraging learners to put down their own ideas is to develop critical thinking by each student. The various parts in the record, which are labeled "own reactions," include: one's own professional situation, relating theory to practice, relating practical suggestions to one or more theoretical considerations, pointing out weaknesses of proposals that have been given, suggesting types of situations under which the proposals might work, and types of situations under which the proposals probably would not work. An attempt is made to urge students to go beyond such superficial comments as, "That looks good," or "I like that," or "it doesn't seem to me that there's anything in that," to expressions of reasons and a critical analysis of *why* the source or student takes a particular point of view.

Periodically, the learner's record is examined by the instructor and by other students in the class. The paramount purposes of such examinations on the part of the instructor are to get a general understanding of what each individual learner is doing, and to give comments and suggestions which will enable him to improve his work as he plans additional activities in the course. The peer evaluation provides helpful suggestions both to those examining a student record and to the student whose record is examined. The written evaluations and self-evaluations of records which students give are sometimes examined by the instructor. Through reciprocal student evaluations of records, the learner gets guided practice in offering help to other students.

Periodically, learners are encouraged to arrive at tentative conclusions with respect to particular subproblems that have been rather thoroughly investigated.

Self-Evaluation

Self-evaluation is conceived as making judgments concerning present and past activities and achievements *so that better plans may be made for future learning*. Thus conceived, self-evaluation is encouraged from the beginning of the course when the learner starts to identify problems of significance to him. Evaluation is a continuous process throughout the course, and at frequent intervals the student is encouraged to look over what he has done and, in the light of such an analysis, to plan future activities more effectively. At times the instructor and other students give guidance to the individual learner. One type of help in self-evaluation is to aid learners in setting up criteria which may be used in evaluating the individual record.

Below is a list of criteria developed by a subgroup in one class. The questions constituting the criteria can serve as one basis for self, peer, and instructor appraisal of student learning as shown in the record.

1. Are comments analytical, and do they show critical thought?
2. Have sufficient theoretical ideas from psychological resources been gathered?
3. Has a good balance of theoretical and applied ideas been maintained?
4. Have theoretical ideas been related to practice?
5. Have the theories underlying suggested practices been sufficiently examined?
6. Are there types of resources which should have been used more? If so, what are they?
7. Is the setup of the record flexible?
8. What do you think should have been done differently?
9. What types of things do you think you should stress more in future study? In what ways are these likely to improve learning activities?
10. Does it appear that your study, as you have developed it, is likely to improve your professional behavior?

In the course, several assumptions about evaluation are made, such as the following:

The primary purpose of evaluation should be to help the learner see where and how he can become more efficient in learning.

Learner goals and purposes in evaluation should be examined frequently and appropriate modifications should be made. Mental health is one of the paramount goals of self-evaluation. In the broad sense, no individual or group is likely long to remain mentally healthy unless efficiency in self-evaluation is developed. The teacher's own self-evaluation can constitute an important basis for teacher and learner improvement. Evaluation of learning activities *in process* is frequently more helpful in improving learning than evaluation of the *end* products of learning, since the former is the real basis for the latter and represents the place where improvement can best take place. Education of the types carried on in the course should not stop when a final grade has been obtained.

Each teacher and each learner should attempt, with appropriate help, to set up his own principles for evaluation, since it usually is, or at least should be, his own responsibility. All concerned should have a share in deciding the purposes and procedures of the evaluation. The student's share should increase as he gains additional facility. Autocratic evaluation basically presents an impossible job for the teacher, because of the number of students. Also, when attempted, it is likely to be uneven in quality, not only because of the magnitude of the job, but also because the learner himself *with sufficient guidance* is in the best position to evaluate his own work. Individualization in education demands democratic evaluation.

Evaluation should not be thought of simply as the judgment of an end product, but rather as a continuing process which always goes on concomitantly with other phases of learning. It is a continuous appraisal and reappraisal for continuous reorientation of goals and activities. Nor should self-evaluation for learning purposes be one which allows students without guidance to do whatever they feel like doing. An important part of self-evaluation is helping the learner build up appropriate criteria to guide him in making various types of decisions. Evaluation should be made up of frequent analyses of past work, and should be oriented toward systematic plans for changes in future work.[2]

Finding and Using Resources.

Students are encouraged to examine different texts, source books, and other resources during the course. One principle to be kept in mind in getting and using resources is that those resources which are appropriate for one learner in a class may be most inappropriate for another learner in the same class. Some factors contributing to this principle are differences in needs, attitudes, skills, and basic abilities, and in what is considered usable material. The instructor has found that many students need much help in finding and using resources. Many students have been trained to depend almost exclusively on instructors and hence must be weaned gradually.

Purposes in using texts differ somewhat from those in conventional courses. Several texts are used rather than one. Each student uses different texts on a reciprocal basis with other students. One purpose is to help them develop the habit of referring to appropriate sources in problem situations. Informal studies of what teachers typically do in a professional problem situation indicate that such a habit is not currently common. A second purpose is to aid learners in practicing the skills necessary in getting help from educational psychology texts. Extensive use is made of tables of contents, introductory paragraphs in chapters, chapter summaries, book indexes, and other key parts of books. A third purpose of use of texts is to give practice in *critically* examining the theoretical statements and practical suggestions in the texts. Initially, students may shy away from this, being loath to be critical of textbooks which they have come to regard almost as sacred cows. An attempt is made to help the learner see that through a critical analysis of the thinking of others his own thinking is sharpened.

---

[2] For additional discussion and illustrations relating to self-evaluation see: Simpson, op. cit., pp. 173–229.

Different texts are exchanged from time to time by individual learners and they attempt through use of the tables of contents, chapter introductions, chapter summaries, and indexes to use those parts of each book which seem to give the most theoretical or applied help on their particular problems.

The resources used by a student should depend upon the problems for study which he has selected, with the help of the teacher, and also should result partially from guidance given by the teacher during the course in the acquisition, selection, and use of resources. The instructor should have available a variety of bibliographies which are frequently used as "starters." Much utilized are *Psychological Abstracts* and *The Education Index.*

Students make considerable use of human resources, such as other students, teachers on the job, other instructors, and, of course, the instructor himself. They are also encouraged to use many other resources such as observations of classrooms in the public schools, and visits to youth groups. An annotated list of readily available films is given to each student and groups of learners are encouraged to view some films in the visual aids laboratory. Some of these films that small groups of students recommend highly are shown to the whole class at a later date.

Furthermore, they are encouraged to get practice in evaluating the ideas obtained from various resources and always to practice giving critical reactions to these ideas. Some instructors contend that, while the objective of such work is desirable, students are not capable of doing it. It is here maintained that competence is developed through guided practice in these activities.

Developing Effective Interpersonal Relations

One of the critical points in our culture seems to relate directly to interpersonal relations weaknesses. We believe in democracy and in some respects we are democratic. However, many school situations are characterized by essentially autocratic practices. A functional educational psychology must face the problems tied up with the question: Do educational and social psychological researches indicate the need for a study and use of democratic interpersonal relations? Most of the evidence seems to suggest an affirmative answer. Therefore, since the course is considered a teaching-learning laboratory, attention is given to ways of studying and using democratic interpersonal relations. In most cases those who are affected by a particular decision should have an increasingly large share in making the judgments which lead to that decision.

Beginning with the first class meeting, an attempt is made to get participation by all students in planning and managing the work of the class. As the course progresses, the instructor attempts gradually to shift more and more social control and responsibility for learning activities from himself to the students. While early assignments are made largely by the instructor, these are of such a nature that the individual learner can adapt them to his needs and abilities. For example, an early assignment, in brief, might be:

> List specific professional problems related to teaching-learning situations which you have encountered as a student or teacher. Supplement this list by studying professional periodicals which take up problems in your field (for example, English teachers or prospective English teachers might use *The English Journal*). After each problem you list, indicate the situation or specific resource which suggested the problem to you.

In one sense all students have the same assignment; however, it is set up in such a way that each learner can adapt it to his background, interests, and needs. If students had had a more desirable type of training in individual responsibility for learning activities, it is quite possible that a more rapid start could be made in giving them primary responsibility for setting up their own learning activities. However, in view of the fact that learners most often come to the class expecting to be told what to do, when to do it, and how to do it, it is necessary, in line with principles of readiness, to work with the learner where he is and get him started off in assuming responsibility gradually.

The following are some of the approaches which have been found useful in establishing effective communication between the students and the instructor and among the different class members: An advisory committee from the class may help the instructor plan the framework for in-class and out-of-class activities. This group may be selected in a variety of ways, any one of which may be the best in a particular situation. Sometimes the advisory group is made up of volunteers. In other classes it has seemed advisable for the instructor to select representatives from various subject-matter or teaching-level groups if the class believes that would be most desirable. In some classes, where small groups within the class are formed early and students have become acquainted, it is possible to have each group select a representative who serves on the advisory committee for a definite length of time. In other classes, where students have known each other, or each student has known several others, it is sometimes desirable to elect representatives. When nominations and elections are used, such questions as these are sometimes raised: What kinds of students should we try to elect? For example, should only the best students be elected or should we elect some who are weak (as determined by self-evaluation, teacher evaluation, or other-person evaluation) in this type of activity so that they can get practice? Students usually decide it is wise to have some of both types. Typically, it is probably desirable to discuss the method of choosing the advisory group so that the students may get some experience in this kind of learning activity.

In large classes, frequent conferences between the instructor and individual students are not possible. As a substitute, a written communication system can be introduced. On a cover sheet of the learner's record, he can note down his problems and difficulties. The teacher can collect a certain number of learner records with attached comment sheets each day or each week. Between class periods the instructor examines them and returns the records with his comments on the cover sheet at the next class period. These may note learning strengths shown by the record, weaknesses in organization, or other criticisms, and may suggest additional approaches or resources. This arrangement permits the teacher to look over the learning record of each student, see what he is doing, and offer constructive suggestions to the learner. It also permits the student to indicate to the instructor what he believes to be his primary difficulties. In addition to regularly scheduled inspections, it is debatable whether it may sometimes be desirable to have unscheduled inspections by the teacher. They permit the teacher to see work in process rather than a record "fancied up" for the purpose, and so encourage the student to get help on a voluntary basis.

Time is occasionally given in helping learners develop in the various roles that a mature group member uses at various times in classes and small groups. Roles include such ones as are involved in giving leadership to an unstructured group. In some classes a bibliography which deals with individual roles in a group and on problems of "growing a group" is given to students. Individual members study various items of the bibliography and also look up other references of a similar sort and then report back either to the class or to small groups in which they may be working.

It must be recognized that the teacher is in a key position as major class leader for guiding group activities toward democratic values and sound procedures. The democratic teacher is able to take upon himself a great variety of roles in the group. In any particular situation, he attempts to assume the role which promises to be of most value to learners. This will sometimes be a good listening role. To promote a situation in which learners can practice different healthy interpersonal roles, it is necessary for the teacher to depart from some roles traditionally played and assume many different roles at different times. For example, the teacher probably should not be thought of as knowing all the answers to the realistic problems that can come up in the course of discussion. The teacher's job, rather, is one of knowing a lot of answers and also of knowing how to help learners develop ways of assuming the responsibility for getting and analyzing possible answers themselves.

Typically, the learner's concept of his responsibility for class operation needs to be changed. The idea is current that the teacher is almost solely responsible for the whole class and for individual and small group success. Added learner responsibility would seem to be essential for significant improvement in large and small group activities as well as in individual learning.

One aspect of interstudent communication is the setting up of groups. These groups may be set up for different purposes. Some of the purposes which have been found significant are:

To answer each other's questions with respect to the operation of the course and the development of individual learning activities.

To offer critical and constructive suggestions on the learning record of other members of the group.

To develop, as a group, various aspects of a particular problem. For example, students in one group might be working on the question: What are some of the ways of diagnosing disciplinary problems and improving them? One student might work on the matter of improving assignment making, another on the ways of getting and using appropriate resources for class activities, and another on the issue of getting democratic learner participation in class activities.

Some consideration has already been given to evaluation. It is important in this connection that both the teacher and the learner give consideration to ways of evaluating their individual behaviors in interpersonal situations and in planning improvement in such behaviors.

Class cooperation is sought not only in developing plans for in-class and out-of-class activities, but also in setting up evaluative criteria which will be used in promoting improvement in the class, and as a basis for marking where marking is either required by the system in which the class is operating or is felt desirable by both the teacher and students. When marking becomes a dual responsibility of the teacher and learners, it is more likely that an attempt will be made to align the basis of marking with the types of activities which will promote profitable learning. This means that the teacher and students must together examine and re-examine the primary purposes of the course and the degree to which these purposes are being achieved.

In the light of such purposes, the teacher and the students can cooperatively set up criteria on which the marks will be based. Such criteria may differ quite significantly from one course to another, depending on its purposes. For example, if it were a course in appreciation of literature, the primary criterion on which marking would be based might be: Has there been improvement in the appreciation of literature defined in behavioral terms during the semester or year and, if so, to what degree? On the other hand, if it is a professional course such as educational psychology or the psychology of learning for teachers, a major criterion might be: To what level of professional development has this learner grown by the end of the period of study—if the mark is designed to indicate level of professional competency.

Summary

The following points represent what the writer believes to be crucial aspects of the course which has just been described.

Continually, attempts are made by the instructor to have each member of the class voluntarily bring his intelligence to bear upon the total problem of utilizing psychological resources in the identification and the solution of significant professional problems—significant particularly to the individual learner. Generally this process involves setting up goals and purposes, deciding on courses of actions, and attempting a continuous evaluation of the work as it proceeds.

The course for an individual learner is not rigidly set up in advance. In effect, the learner produces a custom-tailored set of learning experiences with the guidance of the teacher.

Goals, study procedures, class procedures, and other factors in the management of teaching-learning situations are kept flexible enough to give considerable attention to the development of learner judgment.

Learning for the student consists of a combination of individual, small-group, and whole-class activities which are planned by those most concerned.

The educational psychology course is, itself, used as a teaching-learning laboratory where informal tryouts of some proposed solutions to educational problems can be made.

Received April 23, 1957.

SEX DIFFERENCES AMONG TEACHERS IN THE USE OF THE CRITERIA OF ADJUSTMENT[1]

Harry Beilin and Emmy Werner

University of Minnesota

The criteria employed by teachers in evaluating behavior problems of children has been a matter of much concern since the classic study by Wickman reported in 1928 (**25**). In that study, the attitudes of teachers and "mental hygienists" were contrasted in regard to a list of children's undesirable traits and behavior. The results indicated greater teacher emphasis upon classroom management, learning, and authority problems; clinicians placing greater emphasis upon withdrawing and other "unsocial" personality characteristics.

The report of these findings is regarded as having had a profound effect upon teacher education programs (**6**) with directly measurable changes in teachers' attitudes evident as early as five years after publication of the study (**1**). To say the Wickman study itself was responsible for the reversal of emphasis in teacher education would not do justice to the tenor of the times which the spirit of the report reflected: namely, the growing awareness among many groups of the psychological approach to the understanding of human behavior. This may account for the astonishment of some observers (e.g., **6**) that the Wickman study should have been so widely quoted and uncritically accepted when it embodied some basic flaws.

The limitations of the investigation have been noted as involving three general problems: methodology, the adequacy of the criterion, and the consistency of behavior related to adjustment. The problem of methodology is not independent of the others. The criticisms are in essence these:

1. *Methodology*. The directions given teachers and clinicians

---

[1] This study was supported in part by a research grant (M-690) from the National Institute of Mental Health, and in part by funds from the Institute of Child Welfare, University of Minnesota. This is one of a series of papers of the Nobles County Project, J. E. Anderson and D. B. Harris, Principal Investigators.

were different in the time to respond and the emphasis upon present seriousness of behavior as against seriousness for future adjustment. The differences in directions should affect the comparability of results for both groups.

Replications of the study and related investigations have employed the same directions for teacher and clinician (or other) groups, using the original directions given clinicians (**7, 12, 20**) or those to both groups (**4, 18, 20**). There have been replications (e.g., 25 years later) retaining the original design (**20**) as well as introducing a variety of modifications (**13**) including changes in the time-seriousness orientation (**7, 9**). The results, in the main, indicate that the directions given *S*'s do account to some extent for the differences obtained between groups, but disparities still emerge. With the passage of time, however, there has been a tendency for teachers' criteria to more closely approximate those of clinicians.

2. *The criterion problem.* The second issue concerns the choice of clinicians as a group against which to refer teachers' attitudes toward children. Watson (**23**) has pointed out that there is no reason to suppose clinicians to be "correct" and teachers not, rather than vice versa. Wickman is criticized for not even considering the possibility. Other investigators (**6, 10, 19**) have dealt with this issue although not with the attention it deserves. It has been our thesis (**2**) that the criteria of adjustment are determined, at least in part, by the status and role of the person who evaluates the adjustment of others, and holding to a set of criteria is not inherently more correct for one group than another. The adequacy of criteria (of adjustment) will depend upon some independent standards related to the values and goals of those making use of the criteria. Since teachers may have very different goals from clinicians regarding the behavior of children, it may not be appropriate to impose the clinicians' goals upon the teacher.

3. *Continuity of behavior related to adjustment.* The last major limitation of the original study was its too ready acceptance of a direct relationship of withdrawing behavior in childhood to maladjustment in adulthood (**14, 23, 25**). No adequate data were available to support this assumption in 1928, and one may venture to say no definitive data are available now to substantiate it. As we have observed elsewhere (**3**), a too ready acceptance of this

assumption has tended to hamper success in studies of the prediction of adjustment from childhood to adulthood.

In the present study, we are concerned with the problem of the criterion. In exploring further the relation of role and status to the criteria of adjustment, we are focusing on the effect of sex roles. It has been suggested by Davis and McGinnis (**6**) that the differences between teachers and clinicians may have been due to sex differences (since in 1928 teachers were more likely to have been women; clinicians more likely to have been men[2]).

Procedure

In the present investigation, verbal descriptions from men and women teachers of the behavior of adolescents judged by them to be well or poorly adjusted were obtained. This procedure differs from Wickman's in that teachers are not provided with a set of behavior descriptions but offer their own. The present approach or modifications of it have been employed by others in investigating teacher attitudes (**5, 13, 16, 17, 26**). The present study differs from most all those cited in that the same amount of emphasis is placed upon the criteria of adjustment as on maladjustment.

Thirty-nine teachers (17 women, 22 men) of the high school of a Minnesota community of 8,000 were interviewed as part of a series of studies in the prediction of adjustment. For the purpose of the present investigation, each teacher was asked to nominate the three best and the three poorest adjusted students (regardless of sex) in that school. The selection of high school teachers as evaluators purposefully limited the age range of the persons they evaluated. The teachers were then specifically asked: "Considering the three people whom you have nominated as best adjusted, what is there about them as a group or individually that makes them stand out as best adjusted?" (The same was then asked with regard to the persons nominated as poorest adjusted.)

The responses of the teacher-evaluators were recorded, content analyzed, and the frequencies of responses for each category tabulated according to the sex of the evaluator. Percentages were computed and are presented in Tables 1 and 2. A critical ratio formula

---

[2] Thompson (**22**) reports that women teachers' rankings agree somewhat better with those of child psychologists than do the rankings by men teachers (data however are not presented). Epstein's study (**8**) is also concerned with sex differences in ranking behavior problems of children.

TABLE 1

DIFFERENCES BETWEEN MEN AND WOMEN HIGH SCHOOL TEACHERS IN MENTION OF CRITERIA OF GOOD ADJUSTMENT

Criteria	Female ($N = 17$) %	Male ($N = 22$) %
1.0 *Emotional-personal* (total)	82.3	100.0
1.1 mature, good judgment, sensible, sound	11.7	36.4*
1.2 pleasant, likeable, sweet sunny disposition, good personality, personality	23.5	36.4
1.3 adaptable	23.5	9.1
1.4 general attitude is good, healthy attitude toward marriage, sex, surroundings, other persons	0.0	18.2
1.5 are secure, don't show off, don't complain, not self-conscious	5.8	27.3*
1.6 well-integrated, emotionally stable, well-balanced, calm, not tense, poised	29.4	40.9
1.7 accepts criticism, has insight into self	23.5	9.1
1.8 vivacious, lively, enthusiastic, energetic, glowing	17.6	4.5
1.9 happy	11.7	0.0
2.0 *Character-control* (total)	94.1	72.7*
2.1 good manners, show respect for authority, attentive, courteous, polite, no discipline problems, well-behaved, accept mores of society	29.4	45.5
2.2 takes responsibility, accepts responsibility, assumes responsibility beyond call of duty, trustworthy, dependable, reliable	58.8	45.4
2.3 honest, sincere, straightforward, genuine	5.8	0.0
2.4 are not led, have principles, mind of their own, independent, self-sufficient, real strong individualist	29.4	9.1
2.5 keep themselves decent, have not wandered, good character, don't get into difficulties, moral conduct o.k., tries hard to follow straight and narrow path	0.0	9.1
2.6 humble, modest	23.5	4.5*
2.7 unselfish, altruistic, wants to serve fellow men, do good for others	5.8	0.0
2.8 good at getting things done under trying circumstances, handles difficulties well	5.8	18.2
3.0 *Social-interpersonal* (total)	88.2	77.3
3.1 charming, meets people well, likes other people to be comfortable, ability to carry on conversation, good mixer, friendly, extrovert, fits in well, gets along well, hosts of friends, well liked, popular	70.6	68.2

TABLE 1—*Continued*

Criteria	Female ($N = 17$) %	Male ($N = 22$) %
3.2 leader	35.3	27.3
3.3 cooperative, willing	11.7	13.6
3.4 aggressive, lots of guts with people	0.0	9.1
3.5 participation in extracurricular activities, community activities	23.5	9.1
4.0 *Achievement in general and specific cultural roles* (total)	88.2	72.7
4.1 ambition to make good, wants to get ahead, doesn't let opportunities go by, are goal-oriented, know exactly what they want, planful, lots of determination, drive, initiative, sticks to things and accomplishes them	58.8	36.4
4.2 good students, use their ability, like school, high scholastic standing, get good marks	41.2	18.2
4.3 not necessarily the best or good students but they try	11.7	9.1
4.4 does good job, carries on job very successfully, terrific worker, tends to business, doesn't loaf, works well independently, industrious, efficient at job	17.6	22.7
5.0 *Abilities and interests* (total)	52.9	45.4
5.1 all round person, variety of interests	23.5	18.2
5.2 have ideas, think for themselves, logical thinkers, inventive, creative	11.7	13.6
5.3 bright, alert, capable, superior, brilliant	29.4	13.6
6.0 *Appearance* (total) attractive, very gracious, very fine dresser, carries herself well, neat appearance	5.8	4.5

* Significant at the 10% level of significance.

for independent proportions (**11**, p. 60) was used to determine the significance of the differences in response by sex to each category. The tables, then, indicate the proportion of each teacher group (male and female) that employed each criterion and where significant differences appear. The specific criteria are grouped into general areas: *social-interpersonal*, statements reflecting the relationship between the *S* and others; *emotional-personality*, statements which indicate adequacies or inadequacies in adjustment in which the social reference is lacking or minimal; *character-control*, statements which indicate the extent or adequacy with

TABLE 2

DIFFERENCES BETWEEN MEN AND WOMEN HIGH SCHOOL TEACHERS IN MENTION OF CRITERIA OF POOR ADJUSTMENT

Criteria	Female ($N = 17$) %	Male ($N = 22$) %
1.0 *Emotional-personal* (total)	82.4	81.8
1.1 seeks attention, shows off, wants to be something he isn't, bluff exterior only	23.5	27.3
1.2 persecution complex, suspicious of adults, feels people are picking on him	17.7	27.3
1.3 dissatisfied with things as they are, looking for something they haven't got	5.8	4.5
1.4 emotionally unstable	41.2	18.2
1.4.1 unstable, inconsistent in moods	11.7	0.0
1.4.2 violent temper, bad temper, uncontrolled, purely emotional, hot-headed, high-strung, quick-tempered	5.8	9.1
1.4.3 nervous, can't settle down, restless, moves around, tense, lives in constant fear	23.5	9.1
1.5 resents suggestions from others, reacts poorly to criticism	0.0	18.2
1.6 stands out as different from others, a queer duck a little bit odd, strange	11.7	9.1
1.7 insecure, lacks confidence in self, afraid to say wrong thing, seeks praise, ill at ease, always full of alibi	11.7	22.7
1.8 stutters when excited, one complaint after another	5.9	9.1
1.9 immature, acts like baby, kind of young	11.7	4.5
1.10 unhappy, gets no enjoyment from things average people do	11.7	4.5
1.11 smiles inappropriately, silly sneer on face	5.9	4.5
1.12 makes fool of self, others feel sorry for him	0.0	4.5
1.13 daydreams, reads escapist books	0.0	4.5
1.14 sissy, feminine characteristics, possible sex problems	0.0	4.5
1.15 no spark	5.9	0.0
1.16 sadistic, hurts animals	5.9	0.0
2.0 *Character-control* (total)	82.4	63.6
2.1 unreliable, not dependable, lacks sense of responsibility, untrustworthy, tatler, lies	0.0	27.3*
2.2 hard kid to handle, very trying, lots of disciplinary problems, negativistic, hostile to authority, defiant, cocky, smart alecky, belligerent attitude, in conflict with mores, rebels, poor attitude, chip on shoulder, wise guy, trouble maker, hostile, gets into trouble in school, mischief maker, poor conduct	58.8	31.8*

TABLE 2—*Continued*

Criteria	Female (N = 17) %	Male (N = 22) %
2.3 boisterous, foolishness and joking all the time, conduct themselves poorly in public, loud, noisy, discourteous, coarse no manners	35.3	22.7
2.4 very dependent, not self reliant, can't stand on own two feet, can't make decisions, mamma's boy, conformist, conforms precisely to accepted pattern	5.9	18.2
2.5 runs with wrong crowd, easily led by others in crime	0.0	4.5
2.6 too independent, goes own way, grows up as sees fit	11.7	0.0
3.0 *Social-interpersonal* (total)	70.6	63.6
3.1 lone wolves, withdrawn, unresponsive, don't communicate, shy, timid, very quiet, passive, aloof reserved, not comfortable with people, typical wallflower, mousy, standoffish	35.3	45.5
3.2 lack of adjustment to social groups, not liked, no friends, not accepted in group, not accepted by others, get along poorly, don't attract people, not friendly with people	41.2	35.3
3.3 are uncooperative with peers	0.0	9.1
3.4 self-centered, snobs, envious of others, critical of others, overbearing	17.1	4.5
3.5 overly aggressive, domineering, too bossy, bully	5.9	4.5
4.0 *Lack of achievement* (total)	47.1	63.6
4.1 no goals, no desire to get ahead, no purposefulness, doesn't know what she wants, just rides along, just rolls along, no get up and go, no ambition to get anything done, lazy, no drive for self-improvement, can't buckle down, lack effort, don't apply themselves, never stick to anything	23.5	45.5
4.2 failure at school, contributed little at school, repeating grade, no interest in school, poor student	23.5	27.7
4.3 can't read, can't spell, trouble in expressing himself	11.7	9.1
5.0 *Lack of ability and interest* (total)	41.2	18.2
5.1 lack of ability, low scholastic ability, not bright, slow to learn, doesn't seem to understand anything, not much on the ball	29.4	13.6
5.2 bored, uninterested, dull, inattentive	11.8	9.1
6.0 *Appearance* (total)	17.6	13.6
6.1 physical handicap, long-term illess, physically	11.7	9.1
6.2 not too attractive, slovenly appearance, dirty, personal habits poor, nothing glamorous, dresses inappropriately	5.8	9.1

* Significant of the 10% level of significance.

which the individual has incorporated social controls into his repertory and the traits developed in the process of doing so; *abilities-capacities,* statements referring to *S*'s abilities, capacities, and skills; *achievement,* statements which indicate adequacy or inadequacy with regard to achievement generally or to specific cultural roles, e.g., school, job; *appearance,* statements concerning personal appearance.

Results and Discussion

The reliability of the categorization process was determined for this series of studies employing two psychologically trained raters as classifiers. The percentage agreement for the classification of specific criteria was 85%. Agreement with regard to area classification was 91%.

For six categories (out of 83) there appear significant sex differences in frequency of response. Since the 10% level of significance has been taken as the cutoff point, it is apparent that the number of significantly differentiating items could have occurred by chance alone. The fact that the pattern among the differences is consistent with other results that have been obtained, however, suggests that these particular differences may be meaningful. The most significant observation to be made concerning the results, however, is that male and female teachers are to a high degree consistent in their employment of the criteria of adjustment.

The differences which do emerge suggest that male high school teachers are more inclined to value maturity, good judgment, dependability and trustworthiness. They tend, as well, to characterize the best adjusted as secure persons and lacking in self-consciousness.

Women high school teachers, on the other hand, place greater emphasis on those items classified together as "good character" items, particularly humility and modesty. In addition, and probably most characteristically, women teachers place more stress upon negativism, hostility to authority (the teacher), discipline problems, etc. as criteria of poor adjustment.

As already indicated, in all other respects the criteria which are employed in describing the least and the best adjusted high school students are essentially the same for both men and women teachers. In assessing the relative frequency of the criterion variables (by ranking) for each sex by *area,* some of the findings already indicated receive further support. The most important area mentioned for both good and poor adjustment by women teachers are the "character-control" (essentially, classroom management) items. For the male teachers, however, the "emotional-personality" items receive prime mention. Achievement and ability variables are mentioned quite frequently (as much as 80%) by the high school teachers in characterizing those best or poorest adjusted.

These results tend to lend support to the hypothesis that there may be some sex differences in adjustment criteria and in the direction suggested by Davis and McGinnis, namely, greater emphasis by men on the emotional-personality criteria; greater emphasis by women on the character-classroom management criteria. Considering the possibility these results could have occured by chance alone, however, makes these findings only suggestive.

Summary and Conclusions

Considerable agreement was found between men and women teachers in the criteria used to characterize best and poorest adjusted high school youth.

The criteria which are used significantly more by men teachers are maturity, good judgment, dependability, trustworthiness, lack of self-consciousness, and being secure as a person.

The criteria employed more characteristically by women teachers are humility and modesty. They also place greater stress in regard to poor adjustment upon negativism, hostility to authority, discipline problems, and getting into trouble in school.

In general, the "character-control" items as a group are more often mentioned as criteria by women teachers, "emotional-personality" items as a group by the men teachers.

Intellectual and achievement variables play an important role in the assessments teachers make of the adjustment of young people.

These results are considered in light of a sex difference hypothesis concerning the differences between clinicians and teachers obtained in the Wickman study.

References

1. Bain, W. E. A study of the attitudes of teachers toward behavior problems. *Child Developm.* 1934, **5,** 19-35.

2. Beilin, H. Effect of social (occupational) role and age upon the criteria of mental health. *J. soc. Psychol,* in press.
3. Beilin, H. The prediction of adjustment over a four year interval. *J. clin. Psychol,* in press.
4. Bowles, H. H. A study of nurses attitudes toward behavior problems in children under hospital care. *Child Developm.* 1937, **8,** 282–288.
5. Boynton, P. L., & McGaw, B. H. The characteristics of problem children. *J. juv. Res.* 1934, **18,** 215–22.
6. Davis, E. A., & McGinnis, E. *Parent Education.* Minneapolis: Univer. Minnesota Press, 1939.
7. Ellis, D. B., & Miller, L. W. Teachers' attitudes and child behavior problems. *J. educ. Pschol.,* 1936, **27,** 501–511.
8. Epstein, L. J. An analysis of teachers' judgments of problem children. *J. genet. Psychol.,* 1941, **59,** 101–107.
9. Laycock, S. R. Teachers' reactions to maladjustments of school children. *Brit. J. educ. Psychol.,* 1934, **4,** 11–29.
10. McClethanon, R. H. Teachers and parents study childrens' behaviors. *J. educ. Sociol.,* 1934, **12,** 325–33.
11. McNemar, D. *Psychological Statistics.* New York: John Wiley, 1955.
12. Mitchell, J. C. A study of teachers' and mental hygienists' ratings of certain behavior problems of children. *J. educ. Res.,* 1942, **36,** 292–307.
13. Peck, L. Teacher reports on the problem of unadjusted school children. *J. educ. Psychol.,* 1935, **26,** 123–138.
14. Preston, G. H., & Shepler, W. McL. A study of the problems of normal children. *Amer. J. Orthopsychiat.,* 1931, **1,** 245–256.
15. Schrupp, M. H., & Gjerde, C. M. Teacher growth in attitudes toward behavior problems of children. *J. educ. Psychol.,* 1953, **44,** 203–214.
16. Seidman, J. M., & Knapp, L. B. Teacher likes and dislikes of student behavior and student perceptions of these behaviors. *J. educ. Res.,* 1953, **47,** 143–149.
17. Snyder, L. M. The problem child in Jersey City elementary schools. *J. educ. Sociol.,* 1934, **7,** 343–352.
18. Sparks, J. N. Teachers' attitudes toward the behavior problems of children. *J. educ. Psychol.,* 1952, **43,** 284–291.
19. Stewart, N. Teachers' concepts of "behavior problems" in *Growing points in educational research.* Wash. D. C. Amer. Educ. Res. Ass., 1949.
20. Stouffer, G. A. W., Jr. Behavior problems of children as viewed by teachers and mental hygienists. *Ment. Hyg.,* 1952, **36,** 271–285.
21. Stouffer, G. A. W., Jr. The attitude of secondary school teachers toward certain behavior problems of children. *Sch. Rev.,* 1956, **64,** 358–362.
22. Thompson, C. E. The attitudes of various groups toward problems of behavior. *J. abnorm. soc. Psychol.,* 1940, **35,** 120–125.
23. Watson, G. A critical note on two attitude scales. *Ment. Hyg.,* 1933, **17,** 59–64.
24. Wickman, E. K. *Children's behavior and teachers' attitudes.* New York: Commonwealth Fund, 1928.

25. Witmer, H. L. Childhood personality and parent-child relationships of dementia praecox and manic depressive patients. *Smith Coll. Stud. Soc. Wk.*, 1934, **4**, 287–378.

26. Yourman, J. Children identified by their teachers as problems. *J. educ. Sociol.*, 1932, **5**, 334–43.

Received April 24, 1957.

CHRONOLOGICAL DIVISIONS OF THE LIFE SPAN

Horace B. English

Ohio State University

Most developmental psychologists object to the notion of distinct "stages" of development. Yet it has been almost universally found convenient to describe individuals by such terms as child, adolescent, adult. To find, for the *Dictionary of Psychological and Psychoanalytic Terms,* how these should be defined, I sent a simple open-ended questionnaire to the fellows of the APA Division on Developmental Psychology, asking them to state the chronological limits applicable for these and similar terms. The freedom of expression encouraged did not make for simplicity of tabulation; but, as hoped, a fairly clearcut consensus as to usage was revealed. One hundred and sixteen questionnaires were sent out, 74 were returned.

Table 1 sets forth the limits which best meet the consensus for each named temporal division, and the percentage of replies favoring those limits. In every case there is a mode in the distribution, but in a few cases a secondary mode shows that a divergent principle of categorization is being employed by a minority. In most instances, however, it is the adjacent categories which are next most favored, suggesting that the minority would easily accept the modal usage.

The numerous comments are interesting. "Neonate" is taken as a complete synonym for "Newborn," and many favor its abandonment as pedantic and unnecessary. (Amen to that, and the same for a lot of similar Greek and Latin neologisms!) The "Gang Age" does not belong in the list; it refers to a "stage" in development, if it means anything, which many respondents doubt. There is no consensus as to when it occurs and little desire to pin it down. "Puberty" is also rejected from this list; it is a physiological condition, not an arbitrary division of the life span to be defined by usage or consensus. Unfortunately, I assumed that adulthood and maturity would be used synonymously and did not list the former term. A few respondents, however, seem to need a period of "early

TABLE 1

DIVISIONS OF THE LIFE SPAN

<table>
<tr><th colspan="2">Infancy</th><th colspan="3">Childhood</th></tr>
<tr><td colspan="2">Fr. Birth to 11.9 Mos.
(82%) (46%)</td><td colspan="3">1 Yr. to 11.9 Yrs.
(33% (41%)</td></tr>
<tr><td>Newborn (Neonate) Birth to 0.9 Mo.
(100%) (42%)</td><td></td><td>Early Childhood (young child)
1 Yr. to 5.9 Yrs.
(30%) (50%)</td><td>Mid-Childhood
6 to 9.9 Yrs.
(22%) (25%)</td><td>Late Childhood (pre-adolescence)
10 to 11.9 Yrs.
(35%) (50%)</td></tr>
<tr><th colspan="3">Adolescence</th><th>Maturity</th><th>Old Age</th></tr>
<tr><td colspan="3">12 Yrs. to 20.9 Yrs.
(60%) (42%)</td><td rowspan="2">21 to 65 Yrs.
(32%) (24%)</td><td rowspan="2">65 Up
(47%)</td></tr>
<tr><td>Early Adolescence
12 to 13.9 Yrs.
(50%) (44%)</td><td>Mid-Adolescence
14 to 15.9 Yrs.
(50%) (45%)</td><td>Late Adolescence
16 to 20.9 Yrs.
(35%) (45%)</td></tr>
</table>

adulthood" between adolescence and maturity; they then define the latter as beginning at 30 or even 40. Most, however, are content to begin maturity with the twenties. Perhaps the gerontologists would be inclined to suggest a later age than 65 as the conventional beginning of old age.

The most significant comments, however, were those expressing reluctance to accept a chronological bounding of any of these periods. Quite rightly, they insist that no exact psychological meaning can be given to most of these terms. That, it seems to me, is precisely why a merely formal definition in chronological terms is best. If, for example, one defines adolescence as beginning with puberty, one seems to imply a specific pattern of psychological behaviors "characteristic" of the stage. Such a clearcut stage probably does not exist. Discussions based on such a usage mislead the layman and do not help the professional.

Let us frankly admit that all the distinctions are arbitrary. There is not one of the life periods which could not perfectly well have its limits shifted forward or backward. Is a "newborn" an infant of less than 10 hours, of 10 days, of 14 days, or of what? It doesn't matter which we say except that it is better for us all to say the same thing. There should be strong reasons given if one elects to deviate from the consensus of expert usage here set forth.

Received May 9, 1957.

A COMPARISON OF READING COMPREHENSION SCORES OBTAINED BEFORE AND AFTER A TIME ANNOUNCEMENT

DESMOND L. COOK

Purdue University[1]

One of the most commonly used methods for measuring the rate of reading is to ask the reader to indicate, at a given time signal, the point he has reached in a given reading selection or passage. Rate scores obtained by this method are usually expressed in terms of the number of words read per unit of time, or some similar ratio.

A variation of the above technique is employed on the reading comprehension test given to all new students entering the University of Iowa. This test consists of a series of 10 reading passages, each followed by several choice type questions designed to measure various comprehension skills and abilities. The total number of items is 80, and the time limit for the test is 90 minutes. At the start of the test, examinees are informed that at the end of 45 minutes (or when one-half of the time has passed) they will be asked to indicate on the answer sheet the last item marked. The item indicated serves as a measure of reading rate and is so reported.

The validity of such a procedure may be open to question since such a score would not be a measure of pure reading rate. It is a measure of the time spent not only in reading the passage but in reading and answering the questions following each passage. Thus it is possible for two students to mark the same item but yet have actually completed different amounts of the test. For example, student A may have just completed marking item 42 at the time announcement, but student B has answered item 42 and gone on to read the next passage. Under the procedure of obtaining rate scores, both students would mark item 42 on the answer sheet. Nevertheless, it has been demonstrated to be an adequate measure of the rate at which students read (**1, 8**).

It seems logical to believe that those examinees who had been reading slowly prior to the time announcement would speed up in order to complete the test within the set time limit. Such a speed-up could result in lower comprehension scores for this group due to the fact that not as much time could be given to careful reading of the passages and answering of the questions. The faster readers meanwhile would be further along in the test and would be less concerned about speeding up to finish the test. They would have sufficient time to finish the test, and, consequently, the comprehension score would not be affected to the same degree as for the slow readers. It is even conceivable that the faster readers might improve their scores since more time could be devoted to the remaining items.

---

[1] The data for this study were obtained when the author was Associate Director, Examinations Service, State University of Iowa, Iowa City, Iowa.

The relationship between the rate of reading and comprehension has been the subject of a rather large number of studies. Traxler (**10**), Traxler and Townsend (**11, 12**), Tinker (**9**), and Stroud (**6**) present rather comprehensive summaries evaluating the research on this problem. Stroud reports that the average correlation between speed and comprehension is around .40 when timed tests are used. As Traxler points out, this condition is to be expected if both measures are secured from the same timed test. Using untimed tests, Stroud and Henderson (**7**) found that for elementary school children there was virtually no correlation between rate and quality of reading. Using both timed and untimed tests with college students, Preston and Botel (**4**) report a correlation of .20 for untimed tests and .48 for timed tests, the latter correlation being significant. Usually these studies relating rate to comprehension have been concerned only with the relationship between these two variables and have not reported on the accuracy of comprehension involved.

Specific study of the accuracy of comprehension scores obtained before and after time announcement used to secure a rate score on a timed test appears not to have been investigated. Flanagan (**2**), Anderson and Tinker (**1**), Tinker (**8**), and Robinson and McCollum (**5**) have approached the problem by noting comprehension accuracy in set time limits, but they do not report comparative results. Flanagan studied the relationship between rate and comprehension using twelfth-grade students trained to read at varying rates. Comprehension was noted to decrease slightly as the students went from a slow to a medium rate and much more from the medium to a rapid rate. When the students were divided into three groups on the basis of the slow rate, each group showed about the same decrease in comprehension scores from medium to rapid rates. The results were obtained, however, using scores for the full testing time, and no evidence was reported on relation of accuracy for items attempted. Anderson and Tinker report the results of a study concerned in part with the number of items attempted and scores on these items in a standard time (i.e., part time) for various parts of the Iowa Silent Reading Test. A correlation of .88 was found for the composite score with regard to scores in standard time with items attempted in the same time. The tests were administered individually to 110 college sophomores. No comparison of comprehension accuracy before and after the standard time announcement was reported. Using tests of varying reading difficulty, Tinker found that comprehension accuracy in a standard time varied from no difference to an accuracy of response of 60–70% for the most difficult test. In this study, the standard time limits varied from test to test and ran from 1¾ minutes to 36 minutes with students being allowed unlimited time to finish the tests. Again no results were presented regarding the comparative accuracy of response before and after the standard time announcement. A study of relationship between accuracy of response to number of items attempted on the Paragraph Comprehension part of the Iowa Silent Reading Test for 37 good and 33 poor readers was made by Robinson and McCollum. The groups were selected on the basis of their comprehension scores on the total test. It was found that the good readers were usually superior to the poor readers in both rate and accuracy. Essentially the same results were obtained when the groups were given the Van Wagenen "A" and "Alpha" tests which were untimed. The authors concluded that efficiency in rate of reading was the more influential factor in reading test superiority. As was the case in the other studies cited above, the comprehension accuracy results were reported for items attempted in either a standard or total time, or the accuracy of response on untimed tests, with no comparison of accuracy before and after a standard time announcement on a timed test.

Purpose

The purpose of this investigation is to compare comprehension scores obtained before and after a time announcement for groups of slow and fast readers. The specific hypotheses to be tested are that the mean proportion of items correct before a time announcement is the same as the mean proportion of items correct after a time announcement for each of the groups of readers. This is essentially the null hypothesis, i.e., that the mean difference between scores before and after the time announcement is zero for each of the experimental groups.

It is possible, of course, that any change in performance might be due to factors other than the time announcement. One such factor might be increasing item difficulty as one goes through the test. Another factor which might tend to depress comprehension scores is fatigue. No direct measure of fatigue was obtained. Therefore, its influence can be considered only on a somewhat speculative basis in this study.

Rejection of the hypothesis at the 1% level of confidence for the slow readers but not for the fast readers may be interpreted as evidence in support of a position that the time announcement does have some effect on subsequent comprehension scores. The exploratory nature of the study justifies the setting of such a high level of confidence (**1**).

Procedure

Three groups of 100 answer sheets each were selected from approximately 1,500 answer sheets for the reading comprehension test administered in September 1955 using the percentile values assigned to the rate score as the criterion of selection.

The first group (Group A) consisted of examinees whose rate scores were above the 90th percentile and who had completed the test. The other two groups (B and C) consisted of slow readers. Group B was made up of examinees whose rate scores were below the 10th percentile and who had completed the test. Group C consisted of examinees whose rate scores were below the 15th percentile but who did not complete the test. The original intent was to select Group C from among the examinees in the lowest decile for the rate score, but the lack of a sufficient number of incompleted papers in this area necessitated selection from the area below the 15th percentile in order to secure 100 papers. It was assumed that the individuals in Group C, by not completing the test after the time announcement, either did not speed up or, if they did, they did so to a limited amount. This group was used as a sort of a "control" against which the performance of the slow readers completing the test might be better judged. No corresponding group of unfinished papers was available for the faster readers since most of this group finished the test.

For each examinee in the three groups, the number of items correct before the time announcement was determined and expressed as a proportion of the items attempted in that period. A similar proportion was also determined for items correct of items attempted after the time announcement. A difference score was found for each examinee by subtracting the proportion correct before from the proportion correct after the time announcement. By this procedure, negative differences would indicate that the individual did not do as well (i.e., had a smaller proportion of items correct of items attempted) after the time announcement as before. The median number of items attempted before and after the time announcement was obtained for each group separately. The *t* test for related measures was used to test the significance of the mean difference for each group. No statistical tests of significance were made for differences between the groups.

Results

The results of the study are presented in Table 1. In this table are given the mean proportion of items correct of items attempted before the time announcement for each group, the mean proportion correct of items attempted after the time announcement, and the mean difference for each group. The median number of items attempted before and after the time announcement for each group is also presented in Table 1.

The results for the fast readers completing the test (Group A)

TABLE 1

ITEMS BEFORE AND AFTER TIME ANNOUNCEMENT

	Median Number Attempted		Mean Proportion Correct		
Group	Before Time	After time	Before time	After time	Diff.
A (*N* = 100)	63	17	.59	.55	−.04
B (*N* = 100)	37	43	.53	.39	−.14**
C (*N* = 100)	33	31	.58	.39	−.19**

** *t* significant at 1% level of confidence.

showed that they do read quickly as evidenced by the fact that over one half of the examinees had completed about 79% of the items before the time announcement. For this group, the mean proportions of items correct before and after the time announcement were quite similar. The mean proportion of items correct before the time announcement was .59. After the announcement, the mean proportion of items correct was .55. The average performance after the time announcement was in a negative direction but not statistically significant at the 1% level.

The results for the two groups of slow readers are quite similar. The slow readers completing the test (Group B) tended to speed up as predicted in that 53% of the items were attempted after, as compared to 47% before, the time announcement. The slow readers not completing the test (Group C) apparently were not greatly affected by the time announcement, as they completed 33 out of the 80 items before the time announcement but only 31 of the remaining 47 items after the time announcement. The slow readers completing the test had a mean proportion of correct answers of .53 before and .39 after the time announcement. Slow readers not completing the test had a mean proportion of .58 before and .39 after the time announcement. Both groups showed a quite similar "lack of success" after, as compared to before, the time announcement, the difference being less for slow readers completing the test than for slow readers not completing the test. The difference in mean proportional values before and after the time announcement were statistically significant at the 1% level.

Summary

The purpose of this study was to compare accuracy of comprehension scores secured for slow and fast readers before and after a time announcement used to secure a measure of reading rate. The specific hypotheses tested were that the mean difference between scores before and after the time announcement was zero for each of the experimental groups.

Fast readers completing the test had a mean difference of −.04. Slow readers completing the test had a mean difference of −.14, while slow readers not completing the test had a mean difference of −.19. Statistically significant differences between the proportion of items correct of items attempted before and after a time announcement were obtained for groups of slow readers but not for fast readers.

These results suggest that a time announcement used to obtain a rate score is associated with more lowered comprehension scores after such an announcement than before it for slow readers but not fast readers, as measured by proportion of items correct of items attempted in the two respective time limits.

References

1. Anderson, Verna L., & Tinker, M. A. The speed factor in reading performance. *J. educ. Psychol.*, 1936, **27,** 621–624.
2. Flanagan, J. C. A study of the effect on comprehension of varying speeds of reading. In *Research on the Foundations of American Education*. Wash. D.C.: Amer. Educational Res. Ass., 1939. Pp. 47–50.
3. Lindquist, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton-Mifflin, 1953.
4. Preston, R. C., & Botel, M. Reading comprehension tested under timed and untimed conditions. *Sch. & Soc.*, 1951, **74,** 71.
5. Robinson, F. P., & McCollum, F. H. Reading rate and comprehension accuracy as determinants of reading test scores. *J. educ. Psychol.*, 1934, **25,** 154–157.
6. Stroud, J. B. A critical note on reading. *Psychol. Bull.*, 1942, **37,** 173–178.
7. Stroud, J. B., & Henderson, Margaret. Rate of reading and learning by reading. *J. educ. Psychol.*, 1943, **34,** 193–205.
8. Tinker, M. A. Speed versus comprehension in reading as affected by level of difficulty. *J. educ. Psychol.*, 1939, **30,** 81–94.
9. Tinker, M. A. The relation of speed of comprehension in reading. *Sch. & Soc.*, 1932, **36,** 1958–160.
10. Traxler, A. E. *Ten years of research in reading*, Bull. 32. New York: Educational Records Bureau, 1941.
11. Traxler, A. E., & Townsend, Agatha. *Another five years of research in reading*, Bull. 46. New York: Educational Records Bureau, 1946.
12. Traxler, A. E., & Townsend, Agatha. *Eight more years of research in reading*, Bull. 64. New York: Educational Records Bureau, 1955.

Received May 15, 1957.

UNDER- AND OVER-AGE GRADE-PLACEMENTS AND SCHOOL ACHIEVEMENT

Jack A. Holmes

University of California, Berkeley

and Carmen J. Finley

Sonoma School Districts, California

This study concludes a series of investigations at the elementary school level designed to establish relationships between the degree of over- and under-age grade-placement and relative success in various school subjects. The first two studies, on Grades V and VI, have already been reported (**3**, **4**). The present paper presents data for Grades VII and VIII and then discusses the theoretical significance of the findings.

The hypothesis being tested holds that certain academic disciplines at the elementary level are more important than others in contributing to nonpromotion, promotion, and acceleration.

Samples, Methodology, Tests

For Grade VII, 293 boys and 312 girls and for Grade VIII, 318 boys and 293 girls were drawn from pupils enrolled in 68 elementary school districts in Sonoma County.

Grade-Placement Deviation (GPD) was determined by computing the difference between each child's Actual Grade-Placement (AGP) and the grade he should have been in according to his chronological age. This latter grade is defined as his Chronological-age Grade-Placement (CaGP). Grade-Placement Deviation, then, is expressed as:

$$\text{GPD} = \text{AGP} - \text{CaGP} + \text{K}$$

where K is a constant (5.0) inserted into the equation to give all GPD values a positive sign.

Table 1 gives the frequency distributions of GPD's for each sex in both grades. Inspection reveals that, according to their

TABLE 1

FREQUENCY DISTRIBUTIONS OF ACTUAL GRADE-PLACEMENT DEVIATIONS

Direction of Deviation	GPD Interval	Grade VII Boys f	Grade VII Girls f	Grade VIII Boys f	Grade VIII Girls f
(Under-aged)	2.3 + 2.8[a]	0	1[b]	0	0
(Accelerated)	1.7 + 2.2	0	0	0	0
↑	1.1 + 1.6	4	9	3	6
	.5 + 1.0	49	72	43	70
Normal	− .1 + .4	125	140	124	123
	− .7 − .2	70	59	83	55
	−1.3 − .8	28	16	42	25
	−1.9 − 1.4	12	11	20	8
↓	−2.5 − 2.0	3	3	3	4
(Over-aged)	−3.1 − 2.6	1	1	0	2
(Retarded)	−3.7 − 3.2	1	0	0	0
N		293	312	318	293

[a] The constant K has been subtracted in order to bring these deviations from the mean back into their original form.

[b] Example: This seventh grade girl is accelerated by +2.3 grades; therefore, according to her chronological age she should have been in the (7.1 − 2.3) or 4.8th grade.

chronological ages, the seventh-grade boys *should* have been distributed from Grades 5.7 to 10.4 and the seventh-grade girls should have ranged from 4.8 to 9.7; yet, the entire group was actually in grade 7.1.

Likewise, Table 1 shows that, in accordance with their chronological ages, the boys and girls of the eighth grade should have been enrolled in Grades 6.5 to 10.6 and 6.5 to 11.2, respectively; yet, they were all in Grade 8.1.

The California Achievement Test Battery, Intermediate Form DD, was administered to all seventh- and eighth-grade children during the first three months of the school year as part of the general county testing program in 1952. The tests in this battery are: reading vocabulary, reading comprehension, arithmetic reasoning, arithmetic fundamentals, mechanics of grammar, and spelling. Reliability coefficients range from .85 to .95.

Table 2

Intercorrelations Obtained on 293 Boys in Grade VII

Variable	1 GPD	2 RV	3 RC	4 AR	5 AF	6 MG	7 SP
1. Grade Placement Deviation (GPD)		.450	.346	.318	.259	.364	.397
2. Reading Vocabulary (RV)			.750	.650	.533	.622	.704
3. Reading Comprehension (RC)				.637	.585	.655	.670
4. Arithmetic Reasoning (AR)					.667	.569	.538
5. Arithmetic Fundamentals (AF)						.564	.496
6. Mechanics of Grammar (MG)							.540
7. Spelling (SP)							
Mean	6.98[a]	6.98	6.75	6.78	6.15	6.70	6.83
Standard Deviation	.68	1.60	1.39	1.01	1.04	1.30	1.65

[a] Reconverted: K-Constant Subtracted.

Findings

Grade VII

Table 2 presents the intercorrelations found among the above variables and Grade-Placement Deviation for the boys. The criterion correlates highest with reading vocabulary ($r = .450$) and lowest with arithmetic fundamentals ($r = .259$).

The matrix in Table 2 was submitted to the Wherry-Doolittle Test Selection Method (7) in order to determine which of the test variables made an independent contribution to the variance of the pupils' GPD's. The multiple correlation ($\bar{R}$) rises as the variables, reading vocabulary, spelling, and mechanics of grammar, are successively selected and cumulatively added, but then decreases if any of the other tests are added. Together these three tests yield a multiple $\bar{R}$ of .46, significant at better than the 1% level of confidence.

Table 3 presents the intercorrelations for the girls. The criterion (GPD) correlates highest with reading vocabulary ($r = .380$) and lowest with arithmetic fundamentals ($r = .233$). A Wherry-Doolittle analysis selects only reading vocabulary and spelling

TABLE 3

INTERCORRELATIONS OBTAINED ON 312 GIRLS IN GRADE VII

Variable	1 GPD	2 RV	3 RC	4 AR	5 AF	6 MG	7 SP
1. Grade Placement Deviation (GPD)		.380	.355	.278	.233	.269	.332
2. Reading Vocabulary (RV)			.770	.569	.477	.644	.619
3. Reading Comprehension (RC)				.636	.578	.703	.670
4. Arithmetic Reasoning (AR)					.618	.574	.489
5. Arithmetic Fundamentals (AF)						.572	.492
6. Mechanics of Grammar (MG)							.598
7. Spelling (SP)							
Mean	7.13[a]	7.18	7.23	6.75	6.40	7.30	7.62
Standard Deviation	.68	1.40	1.34	.84	1.16	1.27	1.68

[a] Reconverted: K-Constant Subtracted.

TABLE 4

INTERCORRELATIONS OBTAINED ON 318 BOYS IN GRADE VIII

	Intercorrelations						
Variables	1 GPD	2 RV	3 RC	4 AR	5 AF	6 MG	7 SP
1. Grade Placement Deviation		.434	.440	.430	.327	.381	.439
2. Reading Vocabulary (RV)			.781	.691	.564	.667	.699
3. Reading Comprehension (RC)				.723	.617	.651	.619
4. Arithmetic Reasoning (AR)					.751	.612	.549
5. Arithmetic Fundamentals (AF)						.634	.517
6. Mechanics of Grammar (MG)							.553
7. Spelling (SP)							
Mean	7.88[a]	7.78	7.51	7.75	7.05	7.40	7.58
Standard Deviation	.68	1.56	1.57	1.31	1.46	1.38	1.74

[a] Reconverted: K-Constant subtracted, remainder algebraicly added to 8.1.

TABLE 5

INTERCORRELATIONS OBTAINED ON 293 GIRLS IN GRADE VIII

Variables	Intercorrelations						
	1 GPD	2 RV	3 RC	4 AR	5 AF	6 MG	7 SP
1. Grade Placement Deviation		.348	.365	.302	.298	.279	.327
2. Reading Vocabulary (RV)			.774	.641	.571	.652	.634
3. Reading Comprehension (RC)				.688	.647	.654	.655
4. Arithmetic Reasoning (AR)					.786	.665	.580
5. Arithmetic Fundamentals (AF)						.656	.572
6. Mechanics of Grammar (MG)							.600
7. Spelling (SP)							
Mean	8.11[a]	8.18	8.03	7.85	7.65	8.14	8.67
Standard Deviation	.71	1.49	1.41	1.28	1.62	1.42	1.67

[a] Reconverted: K-constant subtracted, remainder algebraicly added to 8.1.

as variables making independent contributions to the variance of GPD's for seventh-grade girls. Together, the two tests yield an $\bar{R}$ of .393, significant beyond the 1% level.

Grade VIII

Table 4 presents the intercorrelations for the boys. GPD correlates highest with reading comprehension ($r = .440$) and lowest with arithmetic fundamentals ($r = .327$). The multivariate analysis selects reading comprehension, spelling, and arithmetic reasoning. Together, these yield an $\bar{R}$ of .50, significant beyond the 1% level of confidence.

Table 5 presents the intercorrelations for the girls. GPD correlates highest with reading comprehension ($r = .365$) and lowest with mechanics of grammar ($r = .279$). Analysis (7) selects reading comprehension, spelling, and reading vocabulary. Together, these yield an $\bar{R}$ of .38, significant at the 1% level.

Discussion

In order to compare the results obtained in Grades VII and VIII with those found for Grades V and VI already published

TABLE 6

SUMMARY OF SELECTED TESTS AND CUMULATING $\bar{R}$'S FOR GRADES V, VI, VII, AND VIII

Grade	Boys		Girls	
	Tests	Cumulative $\bar{R}$	Tests	Cumulative $\bar{R}$
V	Reading Vocabulary	.248	Spelling	.170
	Spelling	.268[a]	Reading Vocabulary	.183
	Mechanics of Grammar	.275[b]	Arithmetic Fundamentals	.190
VI	Spelling	.365	Reading Vocabulary	.302
	Arithmetic Reasoning	.393	Spelling	.321
	Mechanics of Grammar	.394	Reading Comprehension	.326
VII	Reading Vocabulary	.447	Reading Vocabulary	.377
	Spelling	.458	Spelling	.393
	Mechanics of Grammar	.464		
VIII	Reading Comprehension	.437	Reading Comprehension	.361
	Spelling	.483	Spelling	.375
	Arithmetic Reasoning	.497	Reading Vocabulary	.379

[a] Cumulative $\bar{R}$ for Reading Vocabulary + Spelling.

[b] Cumulative $\bar{R}$ for Reading Vocabulary + Spelling + Mechanics of Grammar.

(**3, 4**), a summary of the selected tests and the cumulative multiple correlations found for both sexes at all four levels is presented in Table 6. Inspection of Table 6 reveals that:

1. Spelling is selected as an important test making a statistically independent contribution to the variance of GPD in all four grades for both boys and girls. The same cannot be said for any other variable in the present battery.
2. Reading vocabulary and spelling are the two most important determiners of GPDs in Grades V, VI, and VII.
3. The cumulative $\bar{R}$ tends to increase from grade to grade.
4. The cumulative $\bar{R}$ is greater for boys than for girls at each grade level.

TABLE 7

RANK-ORDERS FOR SCHOOL SUBJECTS PLAYING MOST TO LEAST IMPORTANT ROLES IN DETERMINING GRADE-PLACEMENT DEVIATIONS

Calculated from College Students' Estimates	Estimated from the $\bar{R}$ Calculations
Grades V, VI, VII	
(N = 75)	(Ns = 979 + 869 + 611)
1. Reading Comprehension	1. Reading Vocabulary
2. Reading Vocabulary	2. Spelling
3. Arithmetic Fundamentals	3. Mechanics of Grammar
4. Arithmetic Reasoning	4. Reading Comprehension
5. Spelling	5. Arithmetic Reasoning
6. Mechanics of Grammar	6. Arithmetic Fundamentals
Grade VIII	
(N = 75)	(N = 611)
1. Reading Comprehension	1. Reading Comprehension
2. Arithmetic Reasoning	2. Spelling
3. Reading Vocabulary	3.5 Arithmetic Reasoning
4. Mechanics of Grammar	3.5 Reading Vocabulary
5. Arithmetic Fundamentals	5. Arithmetic Fundamentals
6. Spelling	6. Mechanics of Grammar

5. Mechanics of grammar is selected as an important subject for boys in grades V, VI, and VII, but this is not so for the girls.

6. Individual differences in achievement in the *quantitative* subjects (arithmetic reasoning and arithmetic fundamentals, in which boys are supposed to excel) appear to play almost no role in determining whether or not a student shall be retarded, promoted, or accelerated in the elementary schools of this system!

7. Individual differences in achievement in the *linguistic* sub-

jects (reading, spelling, and grammar, in which girls are supposed to excel) have a high premium in regard to GPD.

Could these results have been logically predicted?

Perhaps. The fact is, however, they were not! A class of 75 university students enrolled in a teacher-training program was asked to rate these subjects in the order of their importance for determining GPD's in Grades V through VIII.

Table **7** presents the rank orders obtained by preferential analysis of the estimates of these university students and compares them with those obtained from a preferential analysis of the $\bar{R}$'s of the four grades. The university students stress comprehension and relegate spelling to a low order of importance. Their disagreement with the facts is readily apparent in the almost random relationship of the rank orders of the entries in the two columns, especially for Grades V, VI, and VII.

Theoretical Implications

In essence, these studies recall Keys' observations (**5, 6**) that acceleration is rare; indeed, the number of pupils that are accelerated in any one grade represent only a fraction of those retarded. The writers agree with Keys' contention that this does not make for the optimum learning situation for the bright, the average, or the dull. The usual argument for the policy of nonacceleration is that the child should be kept with his chronological-age group. Such a social milieu is supposed to foster both maximum mental health and learning. However, these data clearly show that the desired homogeneity for chronological age (or for that matter, subject-matter achievement) is simply not obtained.

Baker (**1**) and Baker and Traphagen (**2**) concluded from their studies that often bright pupils may fail in arithmetic, and spelling failure may occur at any level. The relationship of their findings to the present study is clear: bright or dull failure in arithmetic is not important for GPD; and spelling failure, which they found occurs at any ability level, also occurs at any grade level. Its occurrence is of great importance for GPD.

The theoretical implications of the present series of investigations concerning school-grade failures seem clear. The schools should come out openly and tell the students and their parents that success in reading comprehension and arithmetic reasoning is not nearly as important as success in reading vocabulary, spelling, and mechanics of grammar for determining retardation, promotion, or acceleration by the time the child is in Grades V, VI, and VII. *Or*, the schools should make it their business to give a more equable distribution to the weights they place on success in each of the subject-matter fields.

As a matter of fact, for neither the boys nor the girls, in Grades V, VI, VII, and VIII, do individual differences in their *combined achievement* in these six subject-matter areas (i.e., reading vocabulary, spelling, mechanics of grammar, reading comprehension, arithmetic reasoning, and arithmetic fundamentals) contribute more than 25% to the variance of the grade-placement deviations within any one class. On what basis, then, one may ask, are students being grouped together in any one grade? On what basis are they being promoted, retarded, or accelerated?

In a scientific age, the apparent unimportance of success in arithmetic is perplexing. This condition is no doubt widespread in the nation's schools, and it is quite possible that this is one of the major reasons why officials of the armed services complain that a great number of men are not well grounded in arithmetic.

If these data are, in fact, at all representative of elementary schools in general, they point up a pressing need for a re-evaluation of the relative importance of the courses offered in the school curriculum.

In an age of interplanetary exploration, the outcome of the race for world leadership may not depend so much upon the ability of a nation to successfully project a satellite into a precalculated orbit, nor even its ability to launch intercontinental missiles, but, as Professor Edward Teller has so clearly stated, upon its ability to simply take and hold the lead in scientific exploration and production. If the results of the above experiment are at all widespread, and there is reason to believe that they are, then it appears that there must be a revamping of our value judgments on what in the curriculum is of the most worth. More than that, knowledge for its own sake, especially in mathematics and science, must be much more enthusiastically supported and *evaluated* as important accomplishments for children in our elementary schools.

References

1. Baker, H. J. *Characteristic differences in bright and dull pupils.* Bloomington, Ill.: Public Sch. Publishing Co., 1927.
2. Baker, H. J., & Traphegen, V. *The diagnosis and treatment of behavior problem children.* New York: Macmillan, 1935.
3. Holmes, J. A., & Finley, Carmen J. Relative importance of curricular areas for grade-placement deviation in grade V. *Calif. J. educ. Res.*, 1955, **6**, 213–218.
4. Finley, Carmen J., & Holmes, J. A. Relative importance of curricular areas for grade-placement deviation in grade VI. *Calif. J. educ. Res.*, 1956, **7**, 200–205.
5. Keys, N. The underage student in high school and college. *Univer. Calif. Publ. Educ.*, 1938, **7**, 145–272.
6. Keys, N. Should we accelerate the bright? *J. except. Child.*, 1942, **8**, 248–254.
7. Wherry, R. J. A new formula for predicting the shrinkage of co-efficient of multiple correlation. *Ann. math. Statist.*, 1931, **2**, 440–457.

Received June 5, 1957.

ITEM DIFFICULTIES BASED ON END SEGMENTS

Edward L. Clark

Northwestern University

In the analysis of test items, there is a strong tradition, originated by Kelley (**3**), for using the proportion of right answers in the top and in the bottom 27% of *S*s when ordered by a criterion. These end groups are used to determine the index of discrimination of the items, and some writers recommend the use of these end groups, neglecting the middle 46%, for determining the difficulty of the items (**1**). This use of the top and the bottom "quarters" of the criterion group for determining the difficulty of test item fails to distinguish between those items which can be answered by only the top 27% and those items which can be answered correctly by as many as the top 73%. Feder (**2**), using the responses of 144 *S*s to the 70 items of a classroom test, has offered evidence in support of determining item difficulty by the end "quarters" of the criterion group. He concluded not only that the difficulty could be determined by the end groups, but also that the usefulness of this method was not related to the discrimination index of the item.

The purpose of the present study was to compare, by correlation, the indices of difficulty as determined by the end quarters of the criterion group with those as determined by the middle two quarters. In order that the effectiveness of the items might be a variable in the study, correlations between difficulties were made for items of five different levels of discrimination.

The data used in this investigation were 1911 item analyses which had each been based on 400 *S*s. These records had been accumulated over a 12-year period as two scholastic aptitude tests were being constructed: the Clark Analogies Test, in parallel forms and based on a very easy vocabulary, and a second test based on a difficult vocabulary. All items were of the multiple-choice type. The criterion most often used in these original analyses was total test scores. At times it was college grades or a mixed criterion of test scores and grades. The poorest items were

TABLE 1

DIFFICULTIES AS DETERMINED BY THE TWO END QUARTERS AND BY THE MIDDLE HALF OF GROUPS

Phi Coef.	No. of item analyses	End groups Mean	End groups *S D*	Middle groups Mean	Middle groups *SD*	*r*	*SD* of all
0–.24	278	115.0	51.94	114.0	55.89	.984	107.38
.25–.39	816	124.0	36.68	122.8	48.83	.977	85.03
.40–.49	497	118.6	29.26	120.9	39.07	.956	67.59
.50–.59	245	113.6	21.15	116.4	22.20	.912	42.42
.60–up	75	107.7	16.30	105.1	27.40	.872	42.38

Note.—The means are the mean number of correct responses to items by 200 *S*s, while the correlation coefficients are between the two determinations of difficulty for each item.

generally analyzed but once; the better items were analyzed with several groups of *S*s. Although five different indices of discrimination were used at different times, the only index used throughout the period was the phi coefficient based on the number of right answers made by the highest 100 and the lowest 100 *S*s. Consequently, for the purposes of this study the phi coefficient was used to classify the items as to level of discrimination.

The first line of data of Table 1 concerns those items which yielded phi coefficients of less than .25, or items which did not discriminate between the highest and lowest *S*s of the criterion groups. In these 278 analyses of items, the end groups of criterion *S*s made right answers in almost the same manner as did the middle groups. The mean and standard deviation of right answers per item were much the same for each group of 200 *S*s, and the correlation between them was $r = .984$. The standard deviation of the number of right answers made by the entire group was 107.4. These items varied greatly in difficulty.

For items of all levels of discrimination, the average number of correct responses is somewhat above 100, out of a possible 200, but for the items of high discrimination level the standard deviations are much lower. Items which discriminate well between strong and weak *S*s not only have an average difficulty near 50% but individualy are much closer to this level of difficulty.

Some importance should be attached to the fact that the standard deviation of the number right per item for the total group of 400 is more than twice as great as the corresponding standard deviations based on the answers of the end groups. The greater relative variability of item difficulties suggests strongly that more information is being given by the entire group than by the ends alone. Since the standard deviations for the middle group were consistently larger than for the ends, the middle half of *S*s gave better evaluations of difficulty for a majority of the items. However, of more concern in this study are the several coefficients of correlation between the difficulties as determined by the end groups and the middle group. While these correlations are very high, they get progressively lower with the increasing discrimination level of the items. In passing, it should be noted that the use of 400 *S*s in each analysis gave rather stable halves and allowed the highest correlations to approach 1.00. The use of 800 *S*s would doubtless have given even higher correlations and the use of only 100 would certainly have given lower. But the important finding in this study is that the use of the end "quarters" alone for determining item difficulties becomes progressively less satisfactory as one deals with increasingly better items.

The use of the proportion of right answers in the end segments of the criterion group for determining the index of discrimination is admittedly an approximation method, but this fact, in this age of card-sorting machinery, seems little justification for not using all the data at hand for determining item difficulties. However, if large numbers of *S*s are used and if *one's items are of little discrimination value*, the item difficulties may quite accurately be determined from the top and the bottom "quarters" of the criterion group.

References

1. Davis, F. B. Item-analysis data; their computation, interpretation, and use in test construction. *Harv. Educ. Pap.*, 1946, No. 2.
2. Feder, D. D. Estimating of test item difficulty by averaging highest and lowest quarter performance compared with total population count. *Educ. psychol. Measmt*, 1947, **7**, 133–134.
3. Kelley, T. L. The selection of upper and lower groups for the validation of test items. *J. educ. Psychol.*, 1939, **30**, 17–24.

Received October 9, 1957.

TABLE 1

DIFFICULTIES AS DETERMINED BY THE TWO END QUARTERS AND BY THE MIDDLE HALF OF GROUPS

Phi Coef.	No. of item analyses	End groups		Middle groups		*r*	*SD* of all
		Mean	*S D*	Mean	*SD*		
0–.24	278	115.0	51.94	114.0	55.89	.984	107.38
.25–.39	816	124.0	36.68	122.8	48.83	.977	85.03
.40–.49	497	118.6	29.26	120.9	39.07	.956	67.59
.50–.59	245	113.6	21.15	116.4	22.20	.912	42.42
.60–up	75	107.7	16.30	105.1	27.40	.872	42.38

Note.—The means are the mean number of correct responses to items by 200 *S*s, while the correlation coefficients are between the two determinations of difficulty for each item.

generally analyzed but once; the better items were analyzed with several groups of *S*s. Although five different indices of discrimination were used at different times, the only index used throughout the period was the phi coefficient based on the number of right answers made by the highest 100 and the lowest 100 *S*s. Consequently, for the purposes of this study the phi coefficient was used to classify the items as to level of discrimination.

The first line of data of Table 1 concerns those items which yielded phi coefficients of less than .25, or items which did not discriminate between the highest and lowest *S*s of the criterion groups. In these 278 analyses of items, the end groups of criterion *S*s made right answers in almost the same manner as did the middle groups. The mean and standard deviation of right answers per item were much the same for each group of 200 *S*s, and the correlation between them was $r = .984$. The standard deviation of the number of right answers made by the entire group was 107.4. These items varied greatly in difficulty.

For items of all levels of discrimination, the average number of correct responses is somewhat above 100, out of a possible 200, but for the items of high discrimination level the standard deviations are much lower. Items which discriminate well between strong and weak *S*s not only have an average difficulty near 50% but individualy are much closer to this level of difficulty.

Some importance should be attached to the fact that the standard deviation of the number right per item for the total group of 400 is more than twice as great as the corresponding standard deviations based on the answers of the end groups. The greater relative variability of item difficulties suggests strongly that more information is being given by the entire group than by the ends alone. Since the standard deviations for the middle group were consistently larger than for the ends, the middle half of *S*s gave better evaluations of difficulty for a majority of the items. However, of more concern in this study are the several coefficients of correlation between the difficulties as determined by the end groups and the middle group. While these correlations are very high, they get progressively lower with the increasing discrimination level of the items. In passing, it should be noted that the use of 400 *S*s in each analysis gave rather stable halves and allowed the highest correlations to approach 1.00. The use of 800 *S*s would doubtless have given even higher correlations and the use of only 100 would certainly have given lower. But the important finding in this study is that the use of the end "quarters" alone for determining item difficulties becomes progressively less satisfactory as one deals with increasingly better items.

The use of the proportion of right answers in the end segments of the criterion group for determining the index of discrimination is admittedly an approximation method, but this fact, in this age of card-sorting machinery, seems little justification for not using all the data at hand for determining item difficulties. However, if large numbers of *S*s are used and *if one's items are of little discrimination value*, the item difficulties may quite accurately be determined from the top and the bottom "quarters" of the criterion group.

References

1. Davis, F. B. Item-analysis data; their computation, interpretation, and use in test construction. *Harv. Educ. Pap.*, 1946, No. 2.
2. Feder, D. D. Estimating of test item difficulty by averaging highest and lowest quarter performance compared with total population count. *Educ. psychol. Measmt*, 1947, **7**, 133–134.
3. Kelley, T. L. The selection of upper and lower groups for the validation of test items. *J. educ. Psychol.*, 1939, **30**, 17–24.

Received October 9, 1957.

THE INTERRELATIONSHIP OF AREA TEST SCORES AND COOPERATIVE GENERAL CULTURE TEST SCORES

JOHN C. WRIGHT

DePauw University

AND BARRON B. SCARBOROUGH

Florida State University

In the fall of 1954, Educational Testing Service (ETS) made available the Area Tests. These tests were "... designed to assess the broad outcomes of education in the liberal arts" (**4**). A description of the tests further stated that "... the tests have been aimed beyond the details of specific courses in an attempt to reach toward those goals of education which are shared generally by the colleges offering instruction in the liberal arts" (**3**). The Cooperative General Culture Test, also available from ETS, was designed "... to provide an indication of the student's general background in five important areas of the college curriculum" (**4**).

In February 1955, the Area Tests were administered at DePauw University in an effort to gain an over-all estimate of student progress in a liberal arts setting. These tests were administered since an estimate was desired of student progress in acquiring a way of thinking and dealing with problems which transcended the mere recall of a volume of relatively isolated academic information. The Cooperative General Culture Test was not used since there was some evidence to indicate it merely attempts to assess the student's knowledge of a wide range of academic, factual information (**2**). The Cooperative General Culture Test is not considered an adequate measure of liberal education (**1**).

Inasmuch as the Cooperative General Culture Test had been administered two years previously to the same group of students, there was a question of the distinctiveness of the two instruments. Therefore, this study was undertaken to see what relationship existed between the scores of students on the Area Tests and the scores of students on the Cooperative General Culture Test at

DePauw University. Seemingly, the two tests were designed for quite different purposes.

Procedure

In February 1955, the Area Tests (Form I) were administered to the senior class of DePauw University ($N = 269$). These tests were returned to ETS for scoring. Scaled scores were furnished for each student. Of the 269 seniors who took the Area Test, 231 were administered the Cooperative General Culture Test in February 1953.

Conditions of administration were comparable in that the same room was used, both tests were part of a university testing program, and the test administrator in each instance had had years of group testing experience.

Product-moment correlations were computed for the scores of each of the subtests of the Cooperative General Test and the related area of the Area Tests. Thus, the scores for Current Social Problems and History and Social Studies were correlated with the score for the area of Social Studies. The scores for Literature and Fine Arts were correlated with the score for the area of Humanities. The scores for Science and Mathematics were correlated with the score for the area of Natural Science. The intercorrelations were obtained for each sex and for the sexes combined. In all instances, raw scores were used for the Cooperative General Culture Test scores and scaled scores used for the Area Tests scores. Raw scores for the Area Tests were not available.

Results

The intercorrelation indices are shown in Table 1.

In general, the correlation coefficients were higher for the women than for the men. None of the coefficients for the women was below .60 while two of these were in the seventies. The lowest single coefficient was .53. This index, which was for the men, was between Mathematics and the area of Natural Science. Both of the coefficients for the area of Humanities for men were .58. All other coefficients for both men and women were in the sixties.

Discussion

The obtained coefficients of interrelationship of the Area Tests and the Cooperative General Culture Tests were, in general, rela-

TABLE 1
CORRELATIONS OF AREA TEST SCORES AND GENERAL CULTURE TEST SCORES

General Culture Test	Area Test	Senior Men ($N = 123$)	Senior Women ($N = 108$)
Current Social Problems	Social Studies	.64	.66
History & Social Studies	Social Studies	.64	.75
Literature	Humanities	.58	.61
Fine Arts	Humanities	.58	.72
Science	Natural Science	.66	.66
Mathematics	Natural Science	.53	.60

tively high. The coefficients were not as high as might be expected for a two-year interval on a test-retest basis for either test.

Many possibilities exist which might account for the obtained coefficients, including: the acquisition of certain facts is a prerequisite to the utilization of other facts, certain factors are not measured but common to each test used (e.g., general intelligence), and/or there is a communality of factors measured. It is possible that DePauw University students constitute an atypical liberal arts sample. Whatever the reason, the evidence suggests further investigation of the distinctiveness of the two instruments.

Summary

The Area Tests and the Cooperative General Culture Tests were designed for somewhat different purposes. The former was designed to assess the broad outcomes of a liberal arts education, whereas the latter was designed to estimate the general background of the student in certain areas of the college curriculum. This study is an estimate of the interrelationship of the subtest scores of the Area Tests and the scores of the Cooperative General Culture Test for DePauw University students. In February 1955, the Area Tests were given to the senior class of DePauw University. The Cooperative General Culture Test had also been administered to the senior class as sophomores in February 1953. Intercorrelations were computed between the comparable areas of each instrument. The obtained coefficients were relatively high. A number of possibilities might account for the indices. However, the evidence seems sufficient to warrant further investigations of the distinctiveness of the two tests.

References

1. Buros, O. K., (Ed.) *The third mental measurement yearbook*, New Brunswick, Rutgers Univer. Press, 1949.
2. Buros, O. K., (Ed.) *The fourth mental measurement yearbook*, New Jersey: Gryphon Press, 1953.
3. Educational Testing Service *Institutional testing program; handbook for deans and examiners*. Princeton, N. J.: Author, 1954–55.
4. Educational Testing Service. *Catalog of cooperative tests*, Princeton, N. J.: Author, 1956.

Received October 9, 1957.

Book Reviews

DANIEL BROWER, & LAWRENCE E. ABT. *Progress in Clinical Psychology*. New York: Grune & Stratton, 1956. Pp. 364.

This volume fulfills the promise made in the first, which presented progress in clinical psychology through 1951. This study includes primarily the years 1952, 1953, and 1954, with several subjects not appearing in the first volume going back as far as 1946.

The introduction deals with clinical psychology from a transactional point of view; Part 2 includes discussions of the Rorschach method, thematic apperception and fantasy tests, and validity of sentence completion tests and human figure drawings. Part 3 discusses therapy in relation to psychoanalysis, client-centered approach, group methods, reading and other subject disabilities, speech disorders, play, music, art, bibliotherapy, psychodrama, hypnosis, and dreams. Part 4 reviews clinical psychology in correctional institutions, rehabilitation of the physically handicapped, and administrative problems in state institutions. Part 5 takes up problems of neuropsychology, statistics and research.

The new subjects not included in Volume 1 are hypnosis, dreams, speech, reading, and some specialized therapeutic techniques.

The very condensed nature of the studies may be seen from the fact that one sentence may have from three to six references, and eight pages of text may have six pages of bibliography. Many statements are references to other summaries. Including all the bibliographies at the ends of chapters, there are 1560 references. The ten subjects with the most extensive bibliographies are the following: dreams (178 references); group psychotherapy (167); rehabilitation of physically handicapped (163); statistics, reading, etc. (each 155); neuropsychology (114); psychoanalytic techniques (75); play, etc. (70); research (64); hypnosis (60). These numbers must not be taken as a measure of anything, but they perhaps suggest the interest of workers in the past few years as represented by the many authors of the present volume—something which might have been considerably different if there had been other writers.

While giving considerable amounts of information, the study primarily refers the advanced student to most of the best in the field. This is not for beginning students, but implies a foundation in the subject. Written clearly in general, it shows some of the ambiguity, paradoxes, and contradictions now existing. Even with these writers we do not quite get away from "faculties" and "fixed tools of thought." Some of the ambiguity in the field of clinical psychology seems, in the language of the psychology of learning, to have transferred.

In general these are scholarly reports by specialists writing in their separate fields of interest and bring the reader up to date (through 1954) in clinical psychology, probably as carefully and accurately as the difficulties of the subject matter permit.

Of special merit are the wide scope of bibliographical material, excellent critical comments, and suggestions for improvement in procedure and for further research, warnings against faulty and unscientific methods, and indications of the gaps that need more attention and filling in.

A good index of subjects completes the volume.

A. S. EDWARDS

The University of Georgia

FRIEDA FROMM-REICHMANN, & J. L. MORENO (Eds.) *Progress in Psychotherapy*, 1956. Grune & Stratton, 1956. Pp. 352.

This is a presentation by about 45 writers and collaborators of the philosophies, work, and progress of the various schools of psychotherapy. It grew out of the "program offered by our new Section at the annual meeting of the American Psychiatric Association in Atlantic City, 1955." Additions were made to include schools which were not heard at the meeting. It is hoped that it will help "to offer a platform for unprejudiced oral and written presentation and comparison of data by representatives of all schools of psychotherapeutic thinking."

Part 1 presents something of the history and philosophy of psychotherapy. Part 2 deals with principles of psychotherapy. Part 3 presents statements by workers in psychotherapy who indicate the nature and methods each of his own school. Part 4 gives summaries of developments in seven foreign countries. Part 5 is a summary of the present status and possible future of psychotherapy.

Part 3, dealing with various schools, includes discussions on the rediscovery of the patient, Adlerian psychotherapy, psychobiologic psychiatry, Jung's concepts, active analytic therapy, existential analysis, psychodrama, the work of Sullivan and Fromm, the views of Horney, communication, biodynamic integration, client centered therapy, multiple therapy, hypnotherapy, countertransference, problems of schizophrenia, group psychotherapy, and prevention and treatment for psychotic children.

An immense number of techniques are indicated; in fact, in his summary, Moreno states that "One hundred and forty-two techniques are mentioned in this book." They may, according to the same author, be classified under seven headings: "physical, pharmacologic, hypnotic, individual, group, action, and technological psychotherapy." The immense sweep and inclusiveness of the subject matter is matched by general excellence of most of the articles. This might be expected since each school in the major presentations is discussed by a selected representative and worker in that particular school. For the foreign countries, Austria, Great Britain, France, Germany, South America, Spain, and Switzerland, the summaries are generally written by one person.

In general, there appear to be very frank and humble statements with clear recognition, at least by many, of how little we know and how far we have to go in understanding psychotherapy. It may appear to many readers that the book contains an amazing collection of unverified concepts and reified hypotheses. Many clear distinctions are made, but we can even find suggestion that psychotherapy and psychoanalysis are indistinguishable. There are many statements of problems, successes and failures; but there are also suggestions of arrogance existing in the field, of unacceptable dogmatisms, and broad generalizations on insufficient data. In the multiplicity of attempts to find foundations for psychotherapy, it is refreshing to find Masserman referring to sound experimental work. In some cases it would seem that experimental psychology and general psychology of the textbooks were treated with thorough disregard if not with contempt. Some formulations seem to be straight from the armchair.

It is pointed out that practice is probably better than theory and definition, and that much of successful therapy lies in the personality and competence of the therapist. The heavier weighting of the volume is on psychoanalytic concepts; rather little recognition is given to interpretation in terms of learning. The experimental psychologist reading this book might like to ask how far astray can we get when imagined entities are substituted for experimental facts? He would not be favorably inclined towards "omnipotent (unconscious) elements;" and he would be inclined to agree that there is too much jargon, too many "high-falutin words," and the danger of too many "parochial attitudes."

Some of these suggestions of adverse criticism have been frankly recognized by some of the writers and properly deprecated. Lack of scientific procedure is recognized and admitted in some of the presentations; but many of the statements are very carefully worked out with due consideration for scientific psychology and for not letting theory and philosophy of psychotherapy get too far out of hand.

It would be difficult to find any book which would better put the reader in touch with the psychotherapy of 1956, or, should we say, with the psychotherapies of 1956. Two indexes are well made and put the reader in touch with most of the details of the book.

A. S. Edwards

The University of Georgia

J. E. Wallace Wallin. *The Odyssey of a Psychologist.* Lyndalia, Delaware: Author. 1955. Pp. 243.

The author, who has more or less restrained himself in more than three hundred publications (books, journal articles, reviews and news articles), now presents what he wants to say about the efforts of half a century in the interest of handicapped children. The book is dedicated to the very many who "rendered the Author Unflagging Support and Encouragement In the Implementation of Services for Children Handicapped in the Race for Life by Accident, Disease, or Inheritance."

It is an autobiography and has two special concerns: first, "to set forth in some detail the nature of the factors that were influential in the making of a special educationist and clinical psychologist—in fact, that proved to be indispensable for the accomplishment of this purpose; second, to provide some insight into the intricate and complex problems of personal and professional adjustment in the field of education that may retard the progress of the work, or even derail it—not because of any flaw in the work itself but because of human weakness—and lead to waste of public funds and blighted hopes and disappointments."

The report includes personal experience in and knowledge about 31 training schools, colleges, and universities in 18 states. It covers also most of the period of the development of the work to which he dedicated his life and to which he probably gave more of pioneering efforts than has any other person. It is the story of his "pioneering experiences in special education, clinical psychology, and mental hygiene."

More or less chronologically, Wallin recounts his experiences and observations in schools and colleges; successes and failures; sources of encouragement and of discouragement; obstacles, both personal and impersonal; the defense of vested rights; progress and collapse of programs; difficulties in getting bills passed in legislatures; the folly of some psychologists; administrative fairness and unfairness; loyal support and strong opposition; untrained psychologists; personal views, both favorable and unfavorable of individuals; attitudes of M.D.s and the improving relations between psychiatrists and clinical psychologists; some of the ills of schools, of extension work, and of publication.

The Odyssey is interesting as a story; it is significant as giving much of the history of the development of clinical psychology and of special classes for handicapped children. Many details found in this volume have probably never been printed before.

The desire of Wallin was not only to give accurate information, but to render assistance to the curing of ills that still continue: "Little hope exists of removing unwholesome conditions that hamper the efficient functioning of public institutions unless the conditions are frankly revealed to the professions concerned and to the public." And he adds: "Fortunately, much in the recital redounds to the credit, honor, and glory of those who have contributed of their might to pioneering in the specific areas in which I have labored."

This is a significant and interesting supplement to the history of psychology, especially in connection with its applications to human improvement. It appears at a time when the work to which he dedicated his life is receiving greatly increased interest, effort, and substantial support.

The book includes almost a complete bibliography of Wallin's publications (Pp. 217–243).

A. S. Edwards

The University of Georgia

GREGORY A. KIMBLE. *Principles of General Psychology.* New York: Ronald Press, 1956. Pp. 400.

This appears to be a very sincere and conscientious effort to present beginning psychology, with great emphasis upon behavior and the extremely objective point of view. Psychology is defined as the "science of behavior." It deals only with "the aspects of behavior that are observable." Experience, mind, and consciousness are excluded from the definition, but introspective reports are acceptable. "It is not the subject's private experience which counts. It is, rather, the report of such experiences which is open to public inspection, verifiable, and testable. Reports of experience made by introspective psychologists are quite as acceptable as any other scientific records."

The presentation is in four parts: 1. Basic Methodology, including general orientation, psychology as a science, graphic and statistical methods, intelligence and aptitude testing, and functions of intelligence; 2. Sensation and Perception: sensory processes in general, vision and audition, and perception; 3. The Modification of Behavior: psychological development, conditioning and simple learning, selective learning, verbal and motor learning, extinction, habit breaking, and forgetting; 4. Behavior dynamics: the primary drives, secondary motivation, frustration, stress and conflict, personality and adjustment, and behavior pathology.

The author at the outset attempts to clarify the meaning and work of psychology as a science and to correct common misconceptions about psychology. The book proceeds to deal with psychology strictly as an objective science, emphasizing the S-R and R-R formulae but not S-O-R. Emphasis is placed upon understanding the place of theory, and of experiment, and of the refinement of experimental control in real science.

There is an appropriate exposition of graphical methods and a brief statement of statistics in which the median is unfortunately defined as "the middle score in a distribution."

The presentation of psychological facts and principles pays thorough respect to experimental work and admits the S-P-R paradigm in connection with perception. At least the word "experience" leaks in for, "Perception refers to more complex experience in which factors such as learning and motivation often play an important role." Of special note are the 211 well-selected illustrations which not only help in clarification but add much of interest to the book, which in general is simply and clearly written.

The student will be interested in the discussions of behavior dynamics and of mental disorders. The author has attempted to give a careful evaluation of Freud's psychoanalysis. But it will be hard for many to classify him with Copernicus and Darwin, and impossible for more to agree that his psychoanalysis was "The forerunner of all of the psychotherapies." There was considerable psychology of motivation before Freud.

The book is to be praised for its many excellent presentations, discussions and warnings to the student. Its double columns, many illustrations, and emphasis upon psychology as an objective science will appeal to many. It gives a well-selected collection for the beginning student who should read some other references for somewhat different treatment of some subjects. Each chapter gives a few suggestions for further reading and a longer list of references. It contains some most recent material not found in many texts. A subject index and a name index conclude the book.

A. S. EDWARDS

The University of Georgia

GEORGE E. GARDNER, (Ed). *Case Studies in Childhood Emotional Disabilities*, Vol. 2. New York: American Orthopsychiatric Association, 1956. Pp. 368.

The foreword of this volume (Vol. 2) indicates the continued hopes that were present in the presentation of the first volume in 1953. These were that the detailed discussions of cases "by selected Association members in the three fields of Child Psychiatry, Clinical Psychology and Psychiatric Social Work" would provide valuable teaching material for these and allied fields. The success of the first volume, requiring a second printing, has led to the publication of the present one.

The cases presented include collaborative treatment of a mother and son; planned return of child to own family; defense mechanisms of a six-year old boy; an acute separation problem; pupils psychologically absent from school; encopresis; the anal phase; brother identification in an adolescent girl; a young delinquent with psychotic mechanisms; ego treatment; a girl with atypical development; the congenitally blind; a hyperactive, destructive boy; adolescent delinquents; and a borderline psychotic child.

The degree of detail and thoroughness of study and report is at least suggested by the fact that there are only 15 reports in a 368 page book. The shortest report is 10 pages and the longest 43 pages in length, the average being a little more than 24. The amount of work in preparing these reports is more than matched by the expenditure of time in therapy (e. g., 200 visits by one boy and 125 by his mother; treatment for more than 4 years). After such intensive work and detailed reporting, the reader may come to feel that he is quite well acquainted with at least some of the cases.

Whether or not the reader is favorably inclined towards the hypotheses, theories and doctrines underlying the therapy, he may well be impressed by the skill of the therapists, the long continued and persistent efforts to give much needed help, the expensiveness of the program, and the value of the results when they are finally attained. None can study this volume without gaining valuable insight into the skills and techniques of "selected" workers in the field.

In general, the reports are simply and clearly written, and, aside from underlying theory, the student should have no trouble in understanding what he reads.

These reports do not deal with drug therapy, but considering its value and the success in teaching of the first volume, the reviewer cannot help but hope that such a book may appear, which, in similar fashion, will report case studies of the combined use of modern drug therapy and psychotherapy. Perhaps this book is the forerunner of such a volume. If drug therapy cuts time and expense, we may still be forced to agree with one of the discussants who remarked on a report of "four years of work with the mother and five years with William": "Possibly the greatest single contribution of research projects of this nature will be to prove to us that intensive treatment of very disturbed personalities is justified."

Page 368 gives the table of contents of Volume 1.

A. S. Edwards

The University of Georgia

Stanley S. Marzolf. *Psychological Diagnosis and Counseling in the Schools*. New York: Henry Holt, 1956. Pp. 401.

"This text has been written in the belief that the personnel worker in education, regardless of where and how he may wish to specialize, should have a single, basic orientation." Emphasis is placed upon guidance. "Inherent in the conception of education as guidance is the recognition of a need for procedures for working with individuals personally, often privately, and with a high regard for those rights of individuality which characterize a free society."

Part 1 deals with the foundations of diagnosis and counseling: the sociological setting for individual methods, the evaluation of behavior and the theory of diagnosis; Part 2 takes up the case study: pertinent information, direct observation and school records, interviewing and physical measurements; Part 3 deals with testing: the characteristics of tests, administering tests, discussions of different kinds of tests, remedial teaching, counseling in theory and in action, recording and reporting.

In general, the book gives a survey of the work and duties of the personnel worker in schools, especially of the school psychologist. The problems are dealt with extensively rather than intensively, and may, in some respects, seem to overlap too much a course in testing. The discussion of testing is more extensive than that on counseling and therapy, and many readers will wish that more space had been given to the latter subjects.

A list of tests indicating "landmarks in the development of testing" is given on page 194. The Stanford-Binet revision (1916) is included but not the Terman-Merrill revision (1937). The latter is mentioned on page 241. Some of the subjects do not seem to be easily found in the index.

The sweep of subject matter seems to be rather broader than the title implies and appears to have crowded out space that might be expected to be given to actual diagnosis and counseling.

On the other hand, the wide sweep and comprehensiveness may do exactly what the author wished, namely, to orient the student in the field as a whole and the problems that are involved. Much interesting and valuable information is given. There are 17 figures and 21 tables. Many valuable suggestions are given in relation to sources of error in the work of the counselor and warnings against pitfalls and lack of proper preparation and training for the work.

The text concludes with a name index and a subject index.

A. S. Edwards

The University of Georgia

GLENN MYERS BLAIR. *Diagnostic and Remedial Teaching.* New York: Macmillan, 1956. Pp. 409.

The title of this book may raise the question, What is diagnostic teaching? The book is better than the title and deals with diagnostic testing and remedial teaching. It includes certain important subjects of special interest to teachers of elementary schools, but which are also forced upon teachers at higher levels.

The text very properly deals with how to find out what the trouble is before trying to remedy it. What are the causes of reading or of other disability, and then what are the best techniques for remedial treatment?

"Remedial teaching" in this book "refers to the two following types of activities: 1. Eliminating ineffective habits and unwholesome attitudes, and reteaching skills which have been incorrectly learned (remedying defects). 2. Teaching for the first time those habits, skills, and attitudes which have never been learned but should have been, and which are needed by pupils (developing increased competence)."

Part 1 considers diagnosis and remedial teaching of reading: the nature and results of remedial programs, determining causes of reading disability, techniques for remedial treatment, remedial programs in elementary and high schools, junior and senior, reading materials and practice exercises. Part 2 deals with arithmetic, spelling, handwriting, and fundamentals of English. Part 3 discusses improving skill in remedial teaching: making case studies and preparing for remedial teaching.

The text is enriched by 25 Tables, 20 Figures, and 16 Plates.

This is a valuable and timely book in a day when increased demand is being made for both diagnosis and remedial teaching. The importance of the book can be better understood when it is known that children are still being sent to clinics; for example, when a boy 13½ years old, is sent to the clinic with the question, "Why can't he learn to read?" an examination shows that his vision is so defective that he cannot recognize a word on a printed page. Teachers who are trained in the materials of this text should be able to detect and to diagnose many cases before the children reach the age of 13 ½ years and will learn much of remedial work which is not sufficiently well known at the present time.

Generally good advice and recommendations are given; classroom teachers can give many tests, but the reviewer must object to the suggestion that diligent study of *Measuring Intelligence* by Terman and Merrill is a substitute for proper training in the use of any of the Binet revisions. Such examinations require thoroughly well-trained examiners, both to give the examination and to interpret results.

Many references are at the bottoms of pages, and useful bibliographies are at the ends of chapters. A well-made index (Pp. 395–409) completes the book. Much valuable material has been gathered together in this volume.

A. S. Edwards

The University of Georgia

PUBLICATIONS RECEIVED

DIDIER ANZIEU. *Le Psychodrame Analytique Chez L'Enfant.* Paris: Presses Universitaires de France, 1956, pp. 183 (paper)

ERIC BERNE. *A Layman's Guide to Psychiatry and Psychoanalysis.* New York: Simon & Schuster, 1957, pp. 320 (paper). $1.50.

FRANK H. BOWLES. *55th Report of the Director, 1956.* New York: College Entrance Examination Board, 1957, pp. 133 (paper). $.50.

JAMES M. BRADFIELD AND H. STEWARD MOREDOCK. *Measurement and Evaluation in Education.* New York: The Macmillan Co., 1957, pp. 509. $5.50.

MARGARET M. CLARK. *Left Handedness: Laterality Characteristics and Their Educational Implications.* London: University of London Press Ltd., 1957, pp. 214.

ROBERT F. DEHAAN AND ROBERT J. HAVIGHURST. *Educating Gifted Children.* Chicago: The University of Chicago Press, 1957, pp. 276. $5.00.

KATHERINE D'EVELYN. *Meeting Children's Emotional Needs: A Guide for Teachers.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1957, pp. 176. $3.75.

RUDOLF DREIKURS. *Psychology in the Classroom.* New York: Harper & Brothers, 1957, pp. 237. $3.75.

LETITIA FAIRFIELD. *Epilepsy: Grand Mal, Petit Mal, Convulsions.* New York: Philosophical Library, 1957, pp. 159. $4.75.

ARDEN N. FRANDSEN. *How Children Learn: An Educational Psychology.* New York: McGraw-Hill Book Co., 1957, pp. 546. $5.50.

CHARLES R. FOSTER. *Guidance for Today's Schools.* Boston: Ginn & Co., 1957, pp. 362. $5.00.

IRA J. GORDON. *The Teacher as a Guidance Worker.* New York: Harper & Brothers, 1956, pp. 350. $4.50.

JOSEPH KATZ, Editor. *Canadian Education Today.* Toronto: McGraw-Hill Co., Inc., 1956, pp. 243. $3.95.

PAUL LAFITTE. *The Person in Psychology: Reality or Abstraction.* New York: Philosophical Library, 1957, pp. 233. $6.00.

WESLEY P. LLOYD. *Student Personnel Services in Japan.* Washington, D.C.: American Council on Education, 1957, pp. 111. $2.00.

BERNARD J. F. LONERGAN, S.J. *Insight: A Study of Human Understanding*. New York: Philosophical Library, 1956, pp. 785. $10.00.

RUTH V. MATHENLY AND MARY TOPALIS. *Psychiatric Nursing*. Second Edition. St. Louis: C. V. Mosby Co., 1957, pp. 259. $3.50.

RAYMOND G. MCCARTHY. *Teen-Agers and Alcohol*. New Haven: Yale Center of Alcohol Studies, 1956, pp. 188. $4.00.

JOHN B. MINER. *Intelligence in the United States*. New York: Springer Publishing Co., Inc., 1957, pp. 180. $4.25.

I. P. PAVLOV. *Experimental Psychology and Other Essays*. New York: Philosophical Library, 1957, pp. 653.

JUSTIN PIKUNAS AND ALLAN P. FARRELL. *Fundamental Child Psychology*. Milwaukee: Bruce Publishing Co., 1957, pp. 259. $3.50.

DOROTHY ROGERS. *Mental Hygiene in Elementary Education*. Boston: Houghton Mifflin Co., 1957, pp. 497. $5.50.

BARUCH SPINOZA. *The Road to Inner Freedom: The Ethics*. New York: Philosophical Library, 1957, pp. 215. $3.00.

RUTH G. STRICKLAND. *The Language Arts in the Elementary School*. Second Edition. Boston: D. C. Heath and Co., 1957, pp. 464. $5.50.

DONALD E. SUPER. *The Psychology of Careers: An Introduction to Vocational Development*. New York: Harper & Bros., 1957, pp. 362. $5.75.

HARRY F. TASHMAN. *Today's Neurotic Family: A Journey into Psychoanalysis*. New York: New York University Press, 1957, pp. 214. $3.75.

HELEN M. THOMPSON. *Swierl: A Plan for Better Reading*. New York: Vintage Press, 1957, pp. 163. $3.50.

MELVIN M. TUMIN. *Segregation and Desegregation: A Digest of Recent Research*. New York: Anti-Defamation League of B'nai B'rith, 1957.

CLARK E. VINCENT. *Readings in Marriage Counseling*. New York: Thomas Y. Crowell Co., 1957, pp. 500. $4.50.

LISTENING: A CASE OF TERMINOLOGICAL CONFUSION[1]

EDNA LUE FURNESS

University of Wyoming

The effect of the auditory mode of presentation on learning and memory has been investigated in various ways since 1894. One problem which existed in the past and which still exists is that of terminology. Is the apperception of the spoken symbol a matter of hearing, of listening, or of "auding" or comprehending aurally? Another problem is that of the processes involved in the aural reception of the spoken symbol. The purpose of this paper is to show the evolution of terminology over the years and to make a differentiation of the processes involved in hearing, listening, and "auding."

At the turn of the century arose a very promising growth of monographic literature devoted solely to the psychophysics of language as such and devoted also to the conscious process involved in the apperception of spoken symbols. Among those studies is one which still remains one of the basic investigations of aural comprehension. It is a study made by the distinguished American educator, William C. Bagley.

In analyzing the apperception of the spoken sentence, Bagley (**3**) considers symbols apart from their meaning and looks at them simply as different forms of stimuli appealing to one or another of the sense departments. He shows that certain psychophysical principles condition the efficiency of perception of symbols. In the typical auditory symbol, the differences in the expressive stimuli are qualitative, temporal, and intensive. In more concrete terms, the symbolic elements differ in that they are either consonants or vowels. Consonants are distinguished by modal differences of complex nature; they are the result of different forms of adjustment of the various parts of the vocal apparatus. Vowels are distinguished by simpler qualitative differences which result from the modifications of the laryngeal clangs. These clangs are brought about by the changes in the form of the pharynx and the buccal cavity. Temporal differences are in the rapidity with which one symbolic element succeeds another symbolic element. Intensive differences are due to the fact that certain symbol elements may be emphasized and that some other symbol elements may be weakened.

---

[1] This article is part of a study, *A Remedial and Developmental Listening Program*, which was made possible by a grant in aid from the University of Wyoming Graduate School Council.

Bagley continues his explanation by stating that the perception of certain letters is sufficient for the production of the verbal (auditory-kinaesthetic) idea, and for the purposes of conceptual "apperception." He answers the question concerning the character of the associative and recognitory processes that underlie the apperception of symbols by stating that, in general, the consciousness accompanying the apperception of auditory symbols is made up of sensational and affective elements. Some of these elements are aroused peripherally; others are aroused centrally in connections which vary in character with different individuals and under different conditions. "These connections," so Bagley says, "are arranged in patterns which change rapidly into one another, and are in general transitory and fleeting" (**3,** p. 125).

Bagley was concerned with the "what" of auditory language. Twenty-eight years later, Paul T. Rankin became interested in the quantification of oral expression among ordinary people. He found that the average adult spends 70% of the total waking time in some form of communication. Of that time a person spends 9% in writing activities, 16% in reading, 30% in speaking, and 45% in listening (**12,** p. 625ff.).

In the 30 years since Rankin's classic attempt to quantify the importance of auditory language, academic interest in the process of decoding spoken symbols has mounted steadily, but the serious question of what to call this form of linguistic behavior still confronts us. Rankin himself was sensitive to the inadequacies of the term "listening," and he used the word with apology and hesitation (**5,** p. 540).

Because of dissatisfaction with the vagueness of the term "listening," writers have coined a striking array of locutions which attempt to denote something more than the mere direction of attention to sound. In use are such terms as: speech reception, thoughtful hearing, getting the meaning from heard words, listening with understanding to spoken language, comprehending oral utterance, interpreting oral expression, inclining the ear cognitively, hearing and understanding, listening as a language art, and aural comprehension (**8**, p. 234).

We are relatively safe from confusion with the term "aural English" so long as we are reading or writing it, but because "oral" and "aural" are homonyms in most of the major American dialects, the safer locution in an oral-aural context is the synonymous term, auditory English. Its abbreviation, "auding," has steadily gained currency in the literature of language education (**5**, p. 540).

The one who first proposed the term "auding," Don Brown, says that he has adopted it to avoid the ambiguities of such terms as "listening," "hearing," and "understanding," as well as to eliminate the awkwardness of such phrases as "getting the meaning from heard words," "listening with understanding to spoken language," or "interpreting vocal expression." As Brown uses the term, he stresses the need for more than mere awareness of perception of sound. For him, a definition of reading provides the key to a definition of "auding." For example, reading may be defined as the gross process of looking at, recognizing, and interpreting written symbols; "auding" may be defined as the gross process of listening to, recognizing, and interpreting spoken symbols (**6**, p. 128). The terms, then, which are most commonly used to indicate the apperception of the spoken language or auditory language are hearing, listening, "auding" or comprehending aurally. In the remainder of this paper the author will attempt to make a differentiation of some of the functions of these three processes.

Granted, the teaching of listening continues to be on the edge of the language arts problem, and it is still commonly assumed that hearing and listening are identical. Unfortunately, the dictionary does not offer a great deal of help in this matter of terminology. Webster says that to *listen* means "to give ear," "to hearken," "to give heed," "to hear with attention," and that to *hear* means "to have the sense or faculty of perceiving sound." Webster states further that to "hear does not necessarily imply, as *hearken* (now only poetical) and *listen* always do, attention or application; e.g., one may *hear* without *listening*, or may *listen* without *hearing*" (**11**, p. 84). Listening comprehension, then, assumes the ability to hear, but hearing does not assume the ability to listen. Hearing and listening are not synonymous. The medical doctor's concern is with hearing; the teacher's concern is with listening (**7**, p. 140).

The words "apprehend" and "comprehend" serve rather well to distinguish between hearing and listening in the sense sought in the foregoing paragraph. Remembering that apprehend means "to become aware of through the senses" and that comprehend means "to embrace or understand a thing in all its compass and extent," perhaps one could say that hearing is the apprehension of sound and that listening is the comprehension of aural symbols (**11**, p. 84).

Obviously, listening is more than hearing (**4**, p. 140). It involves following attentively the thread of a conversation, the development of an idea, the points of an argument. Like reading, it requires comprehension in terms of past experience of the listener and involves critical examination of what is heard. Whenever attention wanders, a portion of what is being presented is lost (**10**, p. 77).

Next we may consider listening and its relation to "auding." In a highly compressed 17-page chapter on "Auding" in the latest of the triennial "Language Arts" issues of the *Review of Educational Research,* Caffrey defines auding as "the process of hearing, listening to, recognizing, and interpreting or comprehending spoken language" (**9**, p. 121). The author goes on to illustrate the distinction between the first two phases, hearing and listening. He cites case studies in which patients exhibit ability to recognize the sounds of running water, crumpling paper, bells, or footsteps but fail to comprehend a word of their native tongue, or respond to their own names (**8**, p. 238). However, "many language specialists prefer to regard hearing and listening, like consciousness, as necessary CONDITIONS rather than as ASPECTS of the auding process, and restrict their concept of auding to recognizing and interpreting hearing language" (**5**, p. 541).

Poetry, *a cappella* choirs, and auto mechanics make special demands upon the ear, but, as Brown says, the relation between listening skill (control of attention to sounds) and "auding" ability (competence in decoding spoken symbols) is far from clear (**5**, p. 541). That is according to Brown. We may look to the research of other investigators. We find that of Frederick B. Agard and Harold B. Dunkel. Their research for the project called the Investigation of the Teaching of a Second Language, sponsored at the University of Chicago, 1944–1947, was used in the United States Army's language training programs. In these programs, as is well known, the approach was predominantly oral and aural. Agard and Dunkel say: "The probability that the determining letters evoke the phonetic letter-sound images corresponding to them, and that these in turn evoke the complete verbal (auditory-kinaesthetic) images has been stated paradoxically as follows: the ability to hear more depends on the need to hear less" (**2**, p. 32). In other words, aural comprehension is dependent to a great degree upon supplementation. Agard and Dunkel's words are (**2**, p. 33): "We understand more than we hear because we fill in the gaps for ourselves without being aware of so doing. . . . It is this ability to supplement which makes the difference between auditory groping and aural comprehension."

In another work, Agard has noted that adequate "reception" or "comprehension" constitutes one-half of the act of speech communication and is composed of at least three elements (**1**, p. 292):

> 1. Auditory apprehension of the spoken utterance and transmission of the acoustic images to the brain for the purpose of physical or linguistic response.
> 2. Knowledge of the meaning to be attached to the acoustic image.
> 3. Awareness of certain facts or assumptions bearing on the subject of discourse but not uttered in the actual speech situations.

The evidence seems to indicate that the term "listening," as it is commonly employed today, is inadequate. The evidence seems to indicate, too, that comprehending aurally or "auding" is the more adequate term; and that "auding" consists of at least six processes: (*a*) hearing, (*b*) listening, (*c*) recognizing spoken language, (*d*) interpreting oral symbols, (*e*) supplementing meaning and knowledge of the symbols, and (*f*) being aware of facts or assumptions not uttered.

References

1. Agard, F. B. Review of Furness test of aural comprehension of Spanish. In O. K. Buros (Ed.). *Third mental measurements yearbook*. New Brunswick, N. J.: Rutgers Univer. Press, 1950.
2. Agard, F. B., & Dunkel, H. B. *An investigation of second-language teaching*. Boston: Ginn, 1948.
3. Bagley, W. C. The apperception of the spoken sentence: a study in the psychology of language. *Amer. J. Psychol.* 1900, **12**, 80–130.
4. Boyd, Gertrude, & Furness, Edna L. *Diagnostic and instructional procedures in the language arts*. Laramie, Wyoming: Curriculum and Res. Center, Univer. Wyoming, 1956.

5. Brown, D. Concepts and practices in teaching aural English. *English J.*, 1946, **35,** 540–546.
6. Brown, D. Teaching aural English. *English J.*, 1950, **39,** 128–136.
7. Brown, J. Construction of a diagnostic test of listening comprehension. *J. exp. Educ.*, 1949, **18,** 139–146.
8. Caffrey, J. G. An introduction to the auding concept. *Education*, 1949, **70,** 234–239.
9. Caffrey, J. G. Auding. *Rev. educ. Res.*, 1955, **25,** 121–138.
10. National Council of Teachers of English, Commission on the English Curriculum. *Language arts for today's children*. New York: Appleton-Century-Crofts, 1954.
11. Nichols, R. G. Listening: questions and problems. *Quart. J. Speech*, 1947, **33,** 83–86.
12. Rankin, P. T. The importance of listening ability. *English J.* 1928, **17,** 623–630.

Received March 28, 1957.

SOME RELATIONSHIPS BETWEEN SOCIOMETRIC POSITION AND LANGUAGE STRUCTURE OF YOUNG CHILDREN[1]

FRED ROSENTHAL

University of California, Berkeley

The scientific study of language has been receiving increasing attention from psychologists and educators as they have become more conscious of the pervasive nature of language in human behavior. In the present investigation, a small aspect of the general field is isolated for study. It is postulated that a relationship exists between an individual's language behavior and his position within a group. More specifically, this study is concerned with the problem of whether any significant differences can be observed in the spoken language used by children of high and low sociometric position.

A survey of the literature in this field suggests that a relationship between language and sociometric position can be demonstrated. This is derived from the following conclusions of various writers:

1. Spoken language as a social phenomenon has been related to other aspects of personality including social adjustment (**1, 2, 4, 9, 17, 18**).
2. There is some indication that individuals with poor language tend to be at the lower end of the scale in personal-social adjustment (**11, 16, 18**).
3. Social behavior as measured by sociometric methods seems to be related to other aspects of personality (**5, 8, 10, 14**).
4. There seems to be a tendency for the individuals of low sociometric status to possess the less desirable personality traits and to be more emotionally disturbed than those individuals who have higher sociometric status (**5, 13, 15, 19**).

General Procedure

Samples of spoken language from a group of second-grade children frequently chosen on a sociometric questionnaire were contrasted with the language samples of "isolated," rarely-chosen

---

[1] This research was carried out under the direction of Drs. Mary Cover Jones, Harold D. Carter, and Read D. Tuddenham of the University of California as partial fulfillment of the requirements for the Ph.D. degree. The reference to the dissertation appears in the bibliography for those readers who desire a fuller treatment of this problem.

Bureau Ednl. Psy. Research
DAVID H[illegible] TRAINING COLLEGE
Dated

children. For the language record a relatively free interview situation with an adult was employed, and each child was encouraged to talk as much as he wished.

The Interview[2]

The interviews were in two parts—in the first, six open-ended questions were asked; in the second, the child was shown six pictures and asked to tell a story about each. Each interview lasted from 15 to 30 minutes. It was recorded on a Gray Audograph machine and later transcribed to typed protocols.

Sociometric Analysis

An analysis of sociometric position was carried out on 11 second-grade classes (358 children) in six elementary schools in Oakland and Berkeley, California.[3] Each child was assessed in terms of his own class; the number of children in these classes varied from 26 to 37 (mean = 33).

In an individual interview, each child was asked to name three classmates with whom he would like to share each of four particular activities. In treating the data, all the votes a child received from his classmates were totaled so as to provide each *S* with a single score. A method developed by Bronfenbrenner (**3**) was then used to convert the raw scores to *T* scores for each class. These *T* scores can be considered comparable for the total group by evaluating them in terms of the probable number of choices which would occur by chance in a group (class) of a given size. Obtaining many more choices than would be expected by chance, given a class of a certain size, indicates high sociometric status; obtaining fewer choices than would be expected by chance indicates low sociometric status.

Children who received fewer choices in their respective classes than would be expected by chance (3% or less of the time) were placed in the "low sociometric" group; those children who received more choices than would be expected by chance (97% or more of the time) were placed in the "high sociometric" group.

---

[2] These questions and instructions are from material made available to the present writer by Dr. Walter Loban of the University of California who developed the interviews.

[3] The same schools from which the language records were obtained.

Sample

The "high" group contained 10 boys and 10 girls, and the "low" group contained 9 girls and 11 boys. In the "high" group, the ages in months ranged from 90 to 97 with the mean equal to 94 months and the standard deviation equal to 2 months; the figures for the "low" group were the same. The Kuhlman-Anderson IQ scores ranged from 92 to 112 in the "low" group with the mean and standard deviation equal to 103 and 6, respectively. In the "high" group, the IQ range was from 92 to 114 with the mean and standard deviation equal to 104 and 6, respectively. The difference between the two groups on IQ was not statistically significant. The variable of socioeconomic status (in terms of father's occupation) was controlled by having equal representation of children of various status levels in each group.

Development of Language Measures

A General Description of Language

In this investigation, language was evaluated in terms of three major characteristics: Structure, Method, and Communicative Value.

Structure. The structural elements of speech are the basic sounds (phonemes) and the connections of sounds (morphenes, words, etc.)—the building blocks out of which all languages derive their form. More specifically, this category involves a description of the materials which an individual uses in his attempt at expression. These structural elements are similar for all persons of a given language group such as English, German, French, etc.

Method. The second general category—the methodological—concerns the way in which the materials are used by a speaker in communication. Differences in style, word choice, rate of speaking, voice quality are all examples of methodological characteristics of speech.

Communicative Value. This category is closely related to the methodological. Furthermore, communicative value can be thought of as characteristic of a purposive aspect in language. In other words, as soon as verbal behavior is subjected to some nonrandom processes or methods, the language can be said to have some function—a function that is the direct expression of the rationale behind the particular methodological characteristics which are being utilized. Thus, if the aim is communication of information, the structure must be handled according to methodological considerations which have been agreed upon by the sender and recipient. For example, when two English speaking individuals wish to exchange information, the sounds, words, sentences, etc., can not be employed in a random fashon but must follow rules which have been agreed upon or accepted (either covertly or overtly) by both if any meaning is to be transmitted.

An Operational Description of Language

In choosing the actual characteristics of language which were to be used in comparing the records of the children, two considerations were involved. Those aspects of speech were chosen which could be measured and evaluated statistically; and an attempt was made to retain meaningful, not merely measurable, features in terms of the three general characteristics discussed above. The following characteristics of language were used in the analysis:

Total Number of Words.

Communication Units. This aspect of language focuses attention on the meaningfulness, the communicative value of a message. Loban[4] in a study of children's speech has developed and used the concept of communication units:

> A word or group of words conveying a meaningful idea and having a quality of completeness about it is a communication unit. (For instance), "I know a boy with red hair." If you leave out "with red hair" you have lost the essential meaning of that sentence. "I know a boy" does not mean the same thing as "I know a boy with red hair." On the other hand, "yes" can be a unit of communication when it is an answer to "Are you going?"

Four numerical scores were derived from this category:

1. The total number communication units (CU) in a given sample of language.

2. The proportion of words used in CU to the total number of words used in a given sample of language. This proportion was defined as follows:

$$P_{CU} = \frac{\text{total of all words used in CU}}{\text{total number of words in sample}}$$

---

[4] This and the following quotes are from an unpublished study by Dr. Walter Loban of the University of California.

3. The mean of the number of words used in CU. This was defined as follows:

$$\mathrm{M_{CU}} = \frac{\text{total of all words in CU}}{\text{total number of CU } (n)}$$

4. The standard deviation of the mean number of words used as CU. This was defined as follows:

$$SD_{\mathrm{CU}} = \sqrt{\frac{\sum(X - \mathrm{M_{CU}})^2}{n}}$$

$$\mathrm{X} = \text{the number of words in one CU}$$

No. 1 is an indication of how many CU's are given in a sample of language; No. 2 is an indication of how much of the total language sample is valuable as communication; No. 3 indicates the average length of the CU's used in a sample of language; and No. 4 indicates how much the individual tends to vary in the lengths of the CU that he uses.

Maze. Mazes are a series of words conveying no meaning; an indication of "noise" in the message. Loban has also developed and discussed this concept giving some examples of mazes from children's speech. The mazes are in parentheses:

1. A short maze at the beginning of a communication unit and deeply related to that communication unit: "(I'm goin) . . . I'm going to beat him up and kick him in the nose."
2. A short maze in the middle of the communication unit and deeply related to that communication unit: "When I was fixin' ready t'go home, mah mother called me up in the house, an' (an', an', an have to) ah have to get mah hair combed."
3. A long maze not immediately related to the communication unit, at least not deeply related: "I saw a hunter program last Sunday (an' he, an' snow time, he had to have lot hu, wah-na when he, uh, not too many dogs, he) . . . and thas all I think of that picture."

In this category four numerical scores were derived in a similar manner and with similar meanings to those described in connection with communication units:

1. The total number of mazes in a given sample of language.
2. The proportion of words used in mazes to the total number of words used in a given sample of language. This was defined as $1 - \mathrm{P_{CU}}$.

3. The mean number of words used in mazes. This was defined as follows:

$$M_m = \frac{\text{total of all words in mazes}}{\text{total number mazes } (n)}$$

4. The standard deviation of the mean number of words used as mazes. This wes defined similarly as above.

Type-Token Ratio. Johnson has discussed this measure of language extensively and described the many forms in which it can be used. He defines the TTR as:

> . . . a measure of vocabulary 'flexibility' or variability, designed to indicate certain aspects of language adequacy. It expresses the ratio of different words (types) to total words (tokens) in a given language sample (**7**, p. 1).

Verb-Adjective Ratio. This measure is obtained by dividing the number of verbs (active construction) by the number of adjectives (passive construction) in a given sample of language (**2**). In this manner, a score of 100 indicates that the number of verbs and adjectives is equal; a score above 100 indicates that there are more verbs than adjectives in the particular language sample; and a score below 100 indicates that there are more adjectives than verbs. Generally, this ratio seems to be an indicator of how active and moving or how passive and inactive a sample of language tends to be (**2, 4**).

The language measures described above can be related to the three characteristics of language which have been postulated by grouping them as follows:

Structure. This category and its measures indicate the number and types of speech characteristics that these language samples contained.

1. The total number of words.
2. The total number of communication units.
3. The total number of mazes.

Method. This category gives a more qualitative description of language, and these measures indicate the variability, movement, complexity, and the richness of the sample of language.

1. The mean number of words in the communication units.
2. The mean number of words in the mazes.
3. The standard deviation of words in the communication units.
4. The standard deviation of words in the mazes.

5. The type-token ratio.
6. The verb-adjective ratio.

Communicative Value. This category consists of the proportion of the total number of words which are used as communication units (and/or $1 - P_{CU}$). This proportion indicates how much of the sample of language is meaningful communication and how much is "nonsense" from the point of view of a recipient.

Analysis

The Language Measures

In order to obtain a relatively standard situation, the analysis of language records was limited to the responses to the six pictures. The exception to this was in the computation of the total number of words which was obtained by counting all the words spoken by the child during the interview. With respect to the other aspects of language, the first 200 words and the last 200 words given in response to the pictures (or smaller segments when a child gave less than 400 words in his stories to the pictures) were analysed seperately, and the correlation of the scores obtained on the two equal segments allowed some estimate of reliability (internal consistency) of scores. A total score in each category was obtained by averaging the two scores or by adding them when interest was in a total number, e.g., the total number communication units, of mazes, of words in communication units and in mazes (in that part of the protocol used for the analysis). In each case the raw numerical scores have been transformed to normalized T scores for ease of statistical manipulation by the procedure outlined by Guilford (**6**).

The following ten aspects of language were drawn from the protocols of the 40 children for analysis. In the remainder of this paper they will be referred to by number:

1. The total number of words.
2. The total number of communication units (CU).
3. The percentage of the total words used in the CU.
4. The mean number of words used in the CU.
5. The standard deviation of the mean number of words used in the CU.
6. The verb-adjective ratio.
7. The mean number of words used in the mazes.

8. The total number of mazes.

9. The standard deviation of the mean number of words used in the mazes.

10. The type-token ratio.

Statistical analysis of these ten language measures on the total group of 40 children indicated the following characteristics:

1. The ten measures were converted to T scores and yielded roughly equal means (51.7 to 55.4; $\overline{X}_M = 52.4$) and standard deviations (9.53 to 11.93; $\overline{X}_{SD} = 10.63$).

2. There seemed to be some tendency for measures two to ten to be internally consistent in these records (excluding measures seven and nine) with Spearman-Brown reliabilities between two parts of the protocols ranging from .47 to .83 (see Table 1).

3. The intercorrelations of the ten measures did not indicate any significant groupings or clusters, nor any excessive degrees of overlap between these measures on the present sample (see Table 1). This indicates that the ten language categories were probably tapping different aspects of language in the present sample.

TABLE 1

PEARSON INTERCORRELATIONS AND RELIABILITIES FOR THE TEN LANGUAGE MEASURES COMPUTED FROM T SCORES

($N = 40$)

Language Measure	2	3	4	5	6	7	8	9	10	R_{II}
1	−26	−17	55**	−02	−07	−33*	−31*	21	−76**	
2		08	15	24	19	24	24	−29	41**	74
3			−06	−03	14	53**	76**	−40**	28	70
4				49**	−09	14	−23	09	−09	59
5					−22	−02	−04	00	17	47
6						11	18	−13	36*	77
7							20	−88**	−45**	15
8								09	29	83
9									−25	43

** Significance at the .01 level.

* Significance at the .05 level.

Note.—Reliability coefficients were obtained with the Spearman-Brown formula from correlations between the scores on the two segments of the protocols of nine of the language measures.

Analysis of Variance

The next step in investigating the present problem was to analyse the variance of the language measures in relation to the independent variable—sociometric status. In the present situation this was accomplished by using the three-way classification scheme presented by McNemar as case XVII (**12,** pp. 332–335); this model, with one modification, seemed to be the one that was most applicable.

The breakdown of the total degrees of freedom was in terms of R rows (individuals), B blocks (sociometric groups), and C columns (language variables). A modification was necessary because the column effect has been taken out when the ten language means had been essentially equalized by the transformation to *T* scores; and therefore, the total degrees of freedom had ten restrictions, instead of one, induced on them. Furthermore, since the individuals from block to block (high and low sociometric status) are relatively independent, i.e., even though the groups are matched for certain variables, there has been no matching by individuals—there was no R by C interaction, no B by R interaction, and no triple interaction. Table 2 summarizes the analysis of variance.

T ratios were also computed between the high and low sociometric groups for each of the ten language measures in order to allow a more intensive analysis of the language characteristics of the two groups. This is summarized in Table 3.

TABLE 2

ANALYSIS OF VARIANCE

Source	*df*	s^2	F	*p*
Individuals (Rows)	(RB − B) 38	122.8 (s_i^2)	18.1 s_b^2/s_i^2	$<.001$
Sociometric Groups (Blocks)	(B − 1) 1	2227.8 (s_b^2)		
Language (Columns)	O			
B × C Interaction	(B − 1) (C − 1) 9	472.8 (s_{bc}^2)	4.7 s_{bc}^2/s_h^2	$<.001$
Residual	B(R − 1) (C − 1) 342	100.1 (s_h^2)		
Total	(RBC − 1) 390			

Table 3

Standard Errors of the Differences Between the Means and t Ratios Between the Means of the High and Low Sociometric Groups on the Ten Language Measures ($df = 38$)

Language Measure	SDM	t	p
1	2.24	0.92	$> .10$
2	2.45	3.92	$< .01 > .001$
3	2.39	2.78	.01
4	2.21	2.42	.02
5	2.38	1.43	$> .10$
6	2.05	4.83	$< .001$
7	2.91	1.68	.10
8	2.45	1.96	$< .10 > .05$
9	2.22	2.21	$< .05 > .02$
10	2.25	2.42	.02

Results

An overall analysis of variances compared the ten measures of language simultaneously between the high and low sociometric groups. The result of this analysis indicated a nonchance difference between the high and low groups. Furthermore, it was discovered that the profiles (B × C interaction) of means for these language measures were not the same for the two extreme sociometric groups.

The significant B × C interaction in the present situation indicates that the ten language measures do not consistently differ from group to group in the same manner—that is, the high group is not superior to the low group in all measures. This interaction effect would lead to the expectations that: (*a*) not all the measures show a difference between the groups; (*b*) the differences obtained are not all of the same magnitude (equally significant); and (*c*) not all the measures differentiate these groups in the same direction, i.e., the high sociometric group having the "superior" language. In the discussion of the t test analysis presented below, this interpretation of the B × C interaction is upheld.

The ten language measures for the high and low sociometric groups were then compared by t ratios in order to obtain a more detailed picture of the differences in the two groups. This analysis indicated:

1. There was no significant difference between high and low sociometric groups in the total number of words given.

2. The high sociometric group gave a significantly smaller total number of communication units than the low sociometric group.

3. The high sociometric group gave a significantly larger percentage of words used as communication units than the low sociometric group.

4. The high sociometric group gave a significantly larger mean number of words used as communication units than the low sociometric group.

5. There was no significant difference between the high and low sociometric groups in the size of standard deviation of the mean number of words used as communication units.

6. The high sociometric group had a significantly greater verb-adjective ratio (more verbs to adjectives) than the low sociometric group.

7. The high sociometric group had a slightly lower (significant at 10%) mean number of words used as mazes than the low sociometric group.

8. The high sociometric group had a slightly smaller number of mazes than the low sociometric group.

9. The low sociometric group had a significantly larger standard deviation of the mean number of words used as mazes than the high sociometric group.

10. The high sociometric group had a significantly higher type-token ratio (more different words relative to the total number of words) than the low sociometric group.[5]

Discussion and Conclusions

Keeping in mind the hazards of generalizing from this small-scale study, the results suggest certain tendencies in regard to the language characteristics of these two groups of second-grade children. The children of high sociometric status tend toward language characterized by a high percentage (of the total words used) of meaningful communication. Their communication is in longer units. They manifest shorter and fewer aspects of "noise" (mazes) in the language, a high usage of verbs relative to adjectives, and a greater variety of words. This group also seems to be somewhat more variable with respect to these aspects of language, especially in regard to the total number of words and the verb-adjective ratio. This latter is only suggested and could be an artifact of the tendency of this group to be higher on the language measures and, therefore, to have more opportunity for variation from zero.

---

[5] It was noted (in the original study and not described in this paper) that two of the variances for these ten measures (1 and 6) for high and low groups are significant. Although this must be viewed tentatively because the variances taken as a group were not significantly different (**16**), it might indicate the following possibilities: 1. The high sociometric group tends to vary more with respect to total number of words given in an interview than the low sociometric group. 2. The high sociometric individuals tend to vary more in their use of verbs relative to adjectives than the low sociometric group. This finding would tend to invalidate the *t* test between the means of the two groups on this measure, but since the difference between the means is highly significant (.001) there is a strong possibility that a nonchance difference remains, albeit at a lower level of significance.

The children of low sociometric status tend toward language characterized by a smaller percentage (of the total words used) of meaningful communication. Their communication is in shorter units. They use a greater number of "noise" (mazes) aspects, longer and more variable mazes, more adjectives relative to verbs, and less variety in the type of words.

There were no differences found for the two groups with respect to the total number of words used and the relative variability in length of communication units.

It is possible to state these conclusions in terms of the threefold analysis outlined earlier. Although it is not assumed that three independent aspects of language exist in terms of the measures used here, nevertheless it is meaningful to conceive of language in this way for purposes of exposition.

Structure. The available evidence does not indicate that high sociometric status is related to the tendency to use a greater quantity of language. The differences found with respect to the total number of communication units and the total number of mazes seem more related to the other aspects of language. In other words, since the groups do not differ in the total number of words used, the findings with respect to CU and mazes might be more indicative of qualitative (method and communicative value) than quantitative differences.

Method. The language of children of high sociometric status seems to be more active and moving and more varied in terms of the types of words employed. The variability in lengths of the mazes seems to be greater in the low sociometric group, but this could be an artifact as described above. Also, the variability of the lengths of the communication units does not seem to differ for the high and low sociometric groups. Even though the individuals in the "high" group tend to give longer communication units in general, they do not vary the length of these units more than the "low" group.

Communicative value. With respect to the aspects of language which seem to be related to communicative value, the findings indicate that the children of high sociometric status are clearly superior to the children of low sociometric status. The language of "high" children is characterized by more communicative value than that of the "low" children.

In general, there was a tendency for children of high sociometric status to be differentiated from children of low sociometric status in terms of language which is more active, variable (word usage), and communicative to another individual. It seems, therefore, that children of this age and of approximately "normal" IQ tended to use an equal number of words regardless of their sociometric status with their peers in the type of interview situation used in this study. However, the more "popular" children use language of superior quality and social usefulness.

It is interesting to speculate about these findings which suggest that sociometric position is positively related to the quality and social usefulness of spoken language. Without making any assumptions about cause and effect, the following tentative "explanation" is advanced for the present results. The work of other investigators which has been mentioned seems to give some support for these views.

When an individual is interacting in a group so that he has high status (being highly chosen), he will presumably have more frequent and more diverse contact with the other members than an "isolated" individual. This in turn would make more communication channels available to the "popular" individual. Thus, the highly chosen individual will have more destinations to which he can send "messages" and a higher probability that he will receive more communication than a less popular person.

In addition, three major considerations seem to arise which favor the expectation that the verbal communication of a highly chosen, popular individual will be superior to that of an isolate.

1. The highly chosen individual has many channels open for outgoing and incoming communication. He has, therefore, many opportunities to make use of verbal communication.

2. Since learning plays an important role in the development of verbal behavior, those individuals who have greater opportunity to send and receive "messages" (more practice) will presumably have more chances to learn successful verbal habits.

3. The highly chosen individuals are also more likely to be rewarded for their attempts at communication by the interpersonal relations which result from this behavior. Also, active intercommunication between individuals will tend to encourage those habits which lead to meaningful transmission and reception—while discouraging those habits that would interfere with this task.

Summary

Twenty second-grade children of "high" sociometric position were contrasted with twenty second-grade children of "low" sociometric position on their language behavior in an interview with an adult. A conceptual schema of language was proposed in terms of structure, method, and communicative value; and the groups were contrasted on the basis of ten statistical aspects of language derived from this schema.

An attempt was made to control IQ, sex, and age; the groups were compared by an analysis of variance and *t* test techniques. The results showed significant differences between "high" and "low" groups on language behavior.

References

1. Balken, E. R., & Masserman, J. H. The language of phantasy: III. The language of the phantasies of patients with conversion hysteria, anxiety state, and obsessive-compulsive neurosis. *J. Psychol.*, 1940, **10**, 75–86.
2. Boder, D. P. The adjective-verb quotient: A contribution to the psychology of language. *Psychol. Rec.*, 1940, **3**, 310–343.
3. Bronfenbrenner, U. Measurement of sociometric status, structure, and development. *Sociom. Monogr.*, No. 6, 1945.
4. Buseman, A. *Die Sprache der Jugend als Ausdruck der Entwicklungsrhythmik.* Jena: Fischer, 1925.
5. Grossman, Beverly, & Wrighter, Joyce. The relationships between selection-rejection and intelligence, social status, and personality among sixth-grade children. *Sociometry*, 1948, **11**, 346–355.
6. Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1950.

7. Johnson, W. Studies in language behavior. I. A program of research. *Psychol. Monogr.*, 1944, **56**, No. 2 (Whole No. 255). Pp. 1–15.
8. Lindzey, G., & Borgatta, E. F. Sociometric measurement. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. I. Cambridge: Addison-Wesley, 1954.
9. Mann, Mary B. Studies in language behavior. II, The quantitative differentiation of samples of written language. *Psychol. Monogr.* 1944, **56**, No. 2 (Whole No. 255). Pp. 41–74.
10. Marks, J. B. Interests, leadership, and sociometric status among adolescents. *Sociometry*, 1954, **17**, 340–349.
11. McCarthy, Dorothea. Language development in children. In L. Carmichael, (Ed.), *Manual of child psychology*. (2nd ed.) New York: Wiley, 1954.
12. McNemar, Q. *Psychological statistics*. (2nd ed.) New York: Wiley, 1955.
13. Northway, Mary L. Outsiders: A study of the personality pattern of children least acceptable to their age mates. *Sociometry*, 1944, **3**, 10–26.
14. Northway, Mary L., Frankel, Esther B., & Potashin. Reva. Personality and sociometric status. *Sociom. Monogr.* No. 11, 1947.
15. Northway, Mary L., & Wigdor, Blossom T. Rorschach patterns related to sociometric status of school children. *Sociometry*, 1947, **10**, 186–199.
16. Rosenthal, F. Some relations between sociometric position and language structure of young children. Unpublished doctoral dissertation, Univer. of Calif., 1956.
17. Sanford, F. H. Speech and personality. *Psychol. Bull.*, 1942, **39**, 811–845.
18. Sanford, F. H. Speech and personality: A comparative case study. *Character & Pers.*, 1942, **10**, 169–198.
19. Scandrette, O. C. Classroom choice status related to scores on components of the California Test of Personality. *J. educ. Res.*, 1953, **47**, 291–296.

Received April 15, 1957.

TRANSFER OF READING TRAINING FROM NONTECHNICAL TO TECHNICAL MATERIAL

MARVIN SCHWARTZ

U. S. Naval School of Aviation Medicine, Pensacola, Florida[1]

A number of recent studies have attempted to evaluate the effectiveness of reading improvement training given college level *S*'s (e.g., **2**). These studies have been principally concerned with demonstrating that reading training results in increased reading rate, increased comprehension, and changes in academic grades. While these results are important, they do not directly answer a typical criticism leveled at reading improvement courses. The science or engineering student or the executive may express concern that, while this training may improve his ability to read a newspaper or novel, it may actually interfere with his ability to read the technical material associated with his occupation. To state it another way: should the materials employed in reading training courses be geared to the course of study of the individual student, i.e., should philosophy majors be trained on philosophical reading and mathematicians on mathematical material? The purpose of the present study was to determine if reading training utilizing relatively fictionalized, nontechnical material would transfer positively to the reading of technical material.

All beginning naval aviation students are exposed to a reading improvement course as part of their pre-flight training. It has been shown (**3**) that this training results in improved reading speed for the material utilized in training. However, since the purpose of pre-flight training is to prepare the student to be an efficient naval aviator, this reading training might better be omitted unless it can be demonstrated to aid the student in the reading of the technical material associated with his training and occupation.

Method

Subjects

The *S*'s were three classes of students in the U. S. Naval School, Pre-flight. The classes contained 27, 30, and 52 students, respectively. Because of academic difficulties in other courses resulting in failure or setback, and drops for physical reasons, 17, 17, and 40 students, respectively, completed the entire experimental procedure. Therefore, 17 from the group of 40 students were chosen, at random, to complete the experimental population. The general population represented by these three classes consists of cadets with a minimum of two years of college education. Each had been previously screened with an intelligence test, a biographical inventory, and a spatial orientation test before being accepted for flight training.

---

[1] The opinions and conclusions expressed are those of the writer and not, necessarily, those of the Department of the Navy.

Reading Training

There are approximately 17.5 hours available for reading training during the pre-flight curriculum. Some seven of these hours are devoted to instruction in the methods and advantages of improving reading ability. The principal techniques utilized for this purpose are lectures and demonstrations with the Harvard series of reading training films. Some seven hours are devoted to actual reading practice, principally with the Reading Rate Controller. The material utilized for this practice varies from fiction and biography to difficult, but nontechnical, essay. Some examples are: *A Son of the Middle Border*, By Hamlin Garland; *The Way of a Transgressor,* by Negley Farson; and *Modern Arms and Free Men*, by Vannevar Bush. While a wide range of difficulty is represented, none of the material can be characterized as being technical in substance. Previous work (**3**) demonstrated that training on this type of material results in an average improvement of 88% in reading speed for this material, and, after 10 weeks of no further instruction, 90% of this improvement is retained. Comprehension, on the other hand, shows an average decrement of approximately 9% with 5% of this decrement remaining after 10 weeks.

Experimental Material

Since the material employed for the reading training is of a nontechnical nature, the object of the present experiment was to utilize as test material relatively technical selections which, while similar to material encountered in pre-flight and subsequent ground school training, would be unfamiliar to the *S*'s. Although they are given an intensive military indoctrination and training in the use of small arms, students passing through the pre-flight curriculum are not given any instruction about naval weapons as employed in the surface fleet. Therefore, the material chosen for the present experiment consisted of three selections from a standard text on that subject, *Naval Weapons and their Uses*, published by the United States Naval Institute. These were about ships' armor, construction of naval guns, and explosives. They contained 1800, 1700, and 1725 words, respectively. The material emphasized the engineering, physics, and chemistry employed in the construction and operation of naval weapons, as well as technical nomenclature. Comprehension tests were constructed to accompany each selection. They contained a mixture of about one-half each of true-false and multiple choice items, 20 per test.

Previous work had demonstrated that reading speed could be measured with these selections with a reliability ranging from .70 to .80. The reliability of the comprehension tests was somewhat lower, .40 to .50, but of sufficient magnitude to assess group trends, as in the present experiment.

Experimental Design

Each class was tested three times: once on each of the three selections, the order of administration to each class being determined by a latin square drawn at random. The essential features of this design are such that it enables independent evaluations of the changes over sessions of the experiment, differences among the test materials, and differences among the three classes. (Differences due to order of administration of the tests were considered to be negligible and are confounded with class differences in the present experiment—cf. [**1**, p. 273] for further explication of the design.)

The initial reading and comprehension tests were given while the students were in the first week of the pre-flight program. This is an indoctrination period and was approximately three days before the reading training was begun. The posttraining test was given on the day following completion of the reading training. The retention test was administered an average of 12 weeks following the posttraining test. (The Christmas leave period caused this to vary among the classes. The minimum time between these tests was 10 weeks.) At the beginning of each test, the students were told simply that this was "a test of your ability to read and understand material of a technical nature—something you will be expected to do throughout your respective naval careers." All tests were administered in a different setting from the one in which the reading training was conducted.

Results

Table 1 presents the mean reading speed and comprehension scores over the course of the experiment. Reading speed for technical material increased 104% between the initial and posttraining tests and showed a 92% improvement when the initial and retention tests were compared. Comprehension of the material read, however, showed a decrement. There was a 7.5% decrement comparing the initial and posttraining tests and a 5% decrement comparing the initial and retention tests. Table 2 presents the summary of the analysis of variance performed on the reading speed measure.

The principal result is the finding of a highly significant source of variance attributable to differences among experimental sessions ($F = 59.29$; 2 and 96 *df*). Applying the *t* test to the indi-

TABLE 1

MEAN READING SPEED AND COMPREHENSION SCORES

	Speed	Comprehension
Initial	193 wpm	73.0%
Posttraining	393 wpm	65.5%
Posttraining—Initial	+104%	−7.5%
Retention	372 wpm	68.0%
Retention—Initial	+ 92%	−5.0%

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE PERFORMED ON READING SPEED

Source of Variance	*df*	Variance	*F*	*p*
Between Subjects:				
Sessions × tests (Classes)	2	618.17	1.59	—
Residual	48	389.48		
Within Subjects:				
Between sessions	2	6,141.52	59.29	<.01
Between tests	2	321.60	3.12	<.05
Sessions × tests	2	176.99	1.72	—
Residual	96	103.06		

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE PERFORMED ON COMPREHENSION

Source of Variance	*df*	Variance	*F*	*p*
Between Subjects:				
Sessions × tests (classes)	2	30.88	3.32	<.05
Residual	48	9.29		
Within Subjects:				
Between sessions	2	30.77	8.45	<.01
Between tests	2	15.06	4.14	<.05
Session × tests	2	2.53	—	—
Residual	96	3.64		

vidual means indicated that the difference between the initial and posttraining sessions was significant beyond the .01 level (t = 8.93; 96 *df*) while the difference between the posttraining and the retention sessions was not significant (t = 1.05; 96 *df*). In addition, the test concerned with ships' armor was somewhat more difficult to read than the other two (t = 2.05; 96 *df*; $p < .05$).

Table 3 presents the summary of the analysis of variance performed on the comprehension scores. Again, the principal result is a highly significant source of variance attributable to differences among experimental sessions (F = 8.45; 2 and 96 *df*). Again, t tests showed that the difference between the initial and posttraining sessions was significant beyond the .01 level (t = 4.06; 96 *df*) while the difference between the posttraining and retention sessions was not significant (t = 1.41; 96 *df*). In addition, the selection about gun construction showed a significantly higher percentage of comprehension than did the other two tests (t = 2.49; 96 *df*; $p < .05$).

Discussion

The present experiment demonstrated that increased reading speed, resulting from a course of training in reading improvement, transfered to the reading of material of a different type from that employed during the training. After being trained on relatively light, nontechnical materials, the students in the present experiment demonstrated over 90% improvement in reading speed for highly technical material. For present purposes, the evaluation of the reading training in terms of the source of improvement is irrelevant. However, we can briefly state what it probably was *not*. It appears unlikely that the improvement is completely ascribable to increased motivation, i.e., simply giving the students a pep-talk on the desirability of reading faster would not have resulted in the degree of improvement demonstrated. Further, since the training is a required part of the pre-flight curriculum for all students, the increase can not be ascribed to the fact that motivated students seeking reading instruction would benefit from *any* course of instruction. More relevantly, whatever the source of the improvement, the fact that it transfered indicates that the course had the desired effect and that the training has benefits beyond the material employed during training. Although a completely definitive test of the present problem would involve reading training on technical material, testing on technical material, and a comparison with the present program, the high reading speed transfer found in this study would suggest that it is not necessary to train readers of technical material utilizing materials from their specialties.

The finding that comprehension of the material read decreased with increased reading speed bears comment. It should be noted that this decrement is probably not a result of negative transfer from the training on nontechnical material, inasmuch as a similar decrement, of the same order of magnitude, was found for the nontechnical material after training (**3**). Thus, the decrement would appear to be due to increased speed of reading alone. It has become axiomatic in this area to say that the fast reader understands and retains more than the slow reader. The present findings would indicate that this is probably only partially true. It may be true in the sense that (*a*) the faster reader covers a greater amount of material, (*b*) less time elapses from the beginning to the end of any given selection, facillitating the integration of the material, and (*c*) the initially faster reader may also be the more intelligent reader. However, the present results indicate that speed, per se, has a deleterious rather than a beneficial effect. Nevertheless, the magnitude of the decrement demonstrated in the present instance is a small price to pay for the comparatively enormous increase in reading speed. Also, this decrement can probably be overcome by stressing that the student should not increase his speed beyond the limits of his ability to comprehend the material.

Summary

The present experiment demonstrated that reading training utilizing nontechnical reading materials for training transfered to the reading of technical material. College level *S*s demonstrated an average of 104% improvement in reading speed for technical material after training on nontechnical material. After 10 weeks, 90% of the improvement was retained. Accompanying this increase was an average decrement of 5–7% in comprehension of the material read. It was concluded that it is not necessary to train readers of technical material utilizing materials from their specialties.

References

1. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
2. McGinnis, Dorothy. Corrective Reading: A means of increasing scholastic attainment at the college level. *J. educ. Psychol.*, 1951, **42**, 166–173.
3. Schwartz, M. An evaluation of the effectiveness of the reading training given in the US Naval School, Pre-flight. *USN Sch. Aviat. Med. Proj. Rep.*, 1957, No. NM 14 02 11 (Rep. No. 1).

Received June 18, 1957.

LEARNING ABILITY AND RETENTION AFTER LONG PERIODS[1]

D. A. Worcester

University of Wisconsin

The writer, in the summer of 1915, learned 20 100-word selections from the prose writing of Matthew Arnold and Thomas Huxley. Twelve of these were learned by the auditory method, i.e., they were read aloud to him until they could be recited, and nine were studied visually until a recitation could be made. The method allowed silent recitations during the learning process. The learner determined the rate of presentation which might vary during the course of learning. An accuracy of at least 95% was required. The time was the total elapsed time from the beginning of learning until the required degree of accuracy was achieved. Altho the goal was to learn to the point of one successful repetition, the above method might result in a higher degree of learning if there were an underestimate of the degree of memorization. One of the selections is here reproduced—one from Arnold.

> The successful future translator of Homer will have, or he cannot succeed, that true sense for his subject, and that disinterested love of it, which are, both of them, so rare in literature, and so precious; he will not be led off by any false scent; he will have an eye for the real matter, and, where he thinks he may find any indication of this, no hint will be too slight for him, no shade will be too fine, no imperfections will turn him aside—he will go before his advisor's thot and help it out with his own.

Five years later, in the summer of 1920, these selections were relearned. During this latter summer these selections and several others were used by the writer as materials to be presented to 12 other persons who memorized them. For those which he had learned previously this constituted overlearning; as to the other selections, while there was no attempt to memorize them, a high degree of familiarity was gained during the process of reading them, perhaps a total of 100 times each, to other persons.

In the summer of 1936, 21 years after the original learning, the writer again studied some of these selections. He relearned 8 of the paragraphs originally learned in 1915 and relearned in 1920. He also learned 8 selections which were used by him in 1920 as materials to be learned by others—which had been, therefore, read again and again but not formally memorized. In addition he learned 4 wholly new selections, 2 by each method, from each of the authors named.

---

[1] This paper was read before the American Psychological Association in New York, September, 1957.

When the results of the 1936 study were reported to the American Association for the Advancement of Science, the student daily paper of the university where the writer was then employed commented upon them under the caption "Senescence No Bar To Learning." Thus encouraged, the writer undertook some more work along this line in the summer of 1956. He relearned 8 of the selections which he had learned in 1915, relearned in 1920, and relearned again in 1936. He also relearned 4 paragraphs which had been learned in 1915, relearned in 1920, but not reviewed in 1936; and 4 which had been learned in 1915, used as materials for others in 1920, but had not been relearned at that time nor reviewed since. And he learned 4 new passages. In each instance half of the selections were done by one method and half by the other.

The 4 new selections learned in July 1956 were relearned in March 1957, a lapse of time of about 8 months.

All of the selections relearned in 1920 employed the same method as that of the original learning. But in 1936 some were relearned by one method and some by the other. Altho learning was more rapid by the visual method, it appeared to make no difference in relearning time which method had been used first. No attention was given in 1956 to the initial mode of presentation.

Each of the paragraphs learned in 1915 seemed definitely familiar to the subject when relearned in 1936, and this was again true in 1956. However, some of them were felt to be much "warmer" than others.

In 1915 and in 1920 the learning exercises were done in the daytime as a part of the regular day's work. The tasks in 1936 and in 1956 were almost entirely done at night and after a full day's work at a busy summer session of the University of Nebraska. The degree to which time of day and fatigue as well as lapse of time and age may have influenced results is not known.

The writer had done almost no memoriter work in many years, and he was especially interested to know if at his present age he could commit material by rote. There was self-assurance that he could, and he felt an eagerness to show that "He was the man he was." Whether it was from overeagerness or from some other cause, there was a definite feeling of tenseness during the study periods of 1956 and 1957 which cannot be recalled as having been present at the earlier times. There was an unwillingness to take a chance—a hesitancy to try to repeat a selection tho it seemed to be learned—but rather an attitude of "I'd better go over it once more." The very feeling of familiarity itself seemed at times to militate against good performance in the case of relearning. The content of the reading was quickly apprehended but many slips made on details—synonyms were used, for example. It is, of course, to be remembered that the increase in age was the same as the lapse of time between learning exercises.

The results of these studies are shown in the accompanying table. In terms of learning new material there seems to be an increase in time required with age. The writer doubts if time of day or fatigue had much, if any, effect. Even the somewhat jittery feeling experienced may itself be a characteristic of age. It may perhaps be noted that the learning times of 1956 were as low or lower than were the average times of those who acted as subjects of the experiment in 1920 whose ages at that time ranged from 21 to 43 (there was one who was 58). Data for the original studies are shown in references (**1, 2**).

It is striking that, altho the visual method continued to be superior to the auditory for this subject, the increase in learning time was greater by that method. A hypothesis concerning this is that reaction time, which commonly increases with age, affects the activities of reading more than those of listening. At age 67 the visual method was markedly superior to the auditory method at age 26.

The data concerning retention show several interesting things. Material relearned by the auditory method showed substantial savings over that originally learned by that method even after long lapses of time and corresponding increases in age. Material originally learned by the visual method actually required slightly more time in 1956, in most cases, for relearning than for learning. On the other hand, one notes the feeling of familiarity experienced as the selections were reread and also the very important fact that

TABLE 1

LEARNING ABILITY AND RETENTION AFTER LONG PERIODS

I Date	Age	N^a	Average Learning Times[b]			New Material % over 1915	
			A	% over 1915	V		
1915	26	21 A 12	13.03				
		V 9			6.29		
1936	47	4 A 2	14.9	14.4			
		V 2			7.25	15.2	
1956	67	4 A 2	16.05	23.2	10.6	68.5	
		V 2					
			Average Relearning Time			% Savings over 1915	
				% Savings over 1915			
1920	31	21 A 12	7.24	44.44			
		V 9			3.41	45.08	
1936	47	8 A 4	5.2	60.			Relearning of material relearned in 1920
		V 4			3.9	38.	
1936	47	8 A 4	5.4	58.6			Material read many times to others but not memorized in 1920
		V 4			3.9	38.	
1956	67	8 A 4	8.9	31.7			Relearning of material relearned in 1920 and again relearned in 1936
		V 4			6.75	−7.3	
1956	67	4 A 2	7.22	44.6			Relearned in 1920. Not reviewed 1936
		V 2			5.38	14.5	
1956	67	4 A 2	7.50	42.6			
		V 2			6.9	−9	Read to many others in 1920 but not formally relearned
1956 total		16 A 8	8.13	37.7			
		V 8			6.57	−4.45	
				Savings over 1956		Savings over 1956	
1957	68	4 A 2	10.				Relearning after 8 months
		V 2		37.2	8.46	20.3	

[a] A = by auditory presentation, V = by visual presentation.

[b] Time in minutes.

this material was relearned in about 38% less time than was required to learn new material. It has been pointed out that the increase in learning time in 1956 was greater for the visual than for the auditory method. It would appear, then, that the seemingly large loss between 1915 and 1956, especially following visual presentation, is a function more of age than of lapse of time per se. It may be said, too, that the original learning by the visual method was accomplished at a considerably greater speed for this subject than that of others who engaged in the early experiment.

In summary, the ability to memorize by rote meaningful prose was by no means lost at age 67 so far as this writer was concerned, tho the rate of learning had decreased. The decrease was more for the visual method of presentation than for the auditory, tho the visual was still the more effective method. Retention of material relearned and reviewed at earlier times was substantial for that relearned by the auditory method. The results for material relearned visually were not so clear. There was probably a good deal of retention, but the factors associated with age seem to be more operative here than in the auditory mode.

It is to be hoped that others for whom early records are available will make comparisons similar to those reported here.

References

1. Worcester, D. A. Memory by visual and by auditory presentation. *J. educ. Psychol.*, 1925, **16**, 18-27.
2. Worcester, D. A. Retention after long periods. *J. educ. Psychol.*, 1923, **14**, 113-114.

Received June 25, 1957.

THE RELATIONSHIP OF CERTAIN PERSONALITY MEASURES TO EXAMINATION PERFORMANCE UNDER STRESS[1]

NEIL A. CARRIER

Southern Illinois University

One of the growing edges of recent psychology has been the study of the relationship of anxiety and behavior. Taylor (**17**), Taylor and Spence (**19**), Montague (**10**), and Lucas (**6**) have reported studies of the performance of *S*s classified as anxious. Other investigators have tried to increase the anxiety of *S*s by manipulating the stressful aspects of a testing situation and then measuring performance (**3, 5, 12, 16**). In general, these studies indicate that anxiety facilitates performance on simple tasks and inhibits performance on more complex activities.

Perhaps more fruitful than the mere demonstration of a relationship between anxiety and performance are attempts to determine the extent to which certain personality characteristics mediate the anxiety-performance relationship. Deese, Lazarus, and Keenan (**1**) reported an interaction between Winne Neuroticism Scale scores and conditions of shock stress under which *S*s learned nonsense syllables. Their findings failed to appear, however, when Lazarus, Deese, and Hamilton (**4**) repeated their procedures using a more complex learning task. High anxiety now apparently inhibited rather than facilitated performance. While several other studies have tried to relate personality to performance under stress (e.g., **11, 18**), the differences reported are generally not of large magnitude. There is a need for further attempts to relate other personality measures to behavior under stress.

Motivation for the present study developed in part from interest in the work by McKeachie and others at the University of Michigan on the effect of anxiety on the examination performance of college students (**8**). While it is generally believed that small amounts of anxiety facilitate many behaviors, an assumption in

---

[1] Adapted from a dissertation submitted to the Doctoral Program in Social Psychology, University of Michigan. Grateful acknowledgement is extended to Dr. W. J. McKeachie for skillful guidance during this research and for helpful reading of an early draft of this manuscript.

McKeachie's work was that most students come to a course examination situation with so much anxiety about the outcome that any additional anxiety is likely to have an adverse effect on their performance. If this is true, it is possible to experimentally manipulate the stressful aspects of an examination situation and measure the corresponding effects on performance. This was done in the present study. The data were analyzed, however, according to the scores the *S*s had obtained on four previously administered personality measures.

Method

The personality measures. The personality variables selected were four that seemed to have some promise of being related to classroom examination performance. They were the SA-S personality inventory developed by Smith and reported elsewhere (**13, 14, 15**), and the need for achievement and need for affiliation measures developed by McClelland and others (**7**). The SA dimension purports to be a measure of "permeability." Persons scoring high in permeability tend to be sensitive to stimuli, impulsive, socially oriented, and low in ego strength. The S dimension ("stability" at the high end and "anxiety" at the low end) is said to reflect free-floating anxiety, nervous tension, and desurgency. The relevance of these measures to examination performance under stress should be evident. If the highly permeable student is placed in a testing situation which carries a large potential threat, his sensitivity may produce a level of anxiety which will cause a greater performance decrement than that of the relatively impermeable student. And when stress is introduced, the low stability student may show a greater performance decrement than that displayed by the more stable student.

The n Ach variable was used in view of the suggestion (**7**, pp. 110–111, 215, 271) that in males it discriminates between success-oriented (high n Ach) and failure-fearing (moderate n Ach) persons. These differently oriented persons may differ both in the kinds of attitudes they bring to a testing situation and in their abilities to handle the stressful aspects of it. The n Aff variable was used because of the implication that high n Aff persons are made more anxious by being deprived of their affiliative ties than low n Aff persons. If the high n Aff student were removed from some of his fellow students, from his teacher, and from his usual classroom and required to take an important examination in a strange room with a strange "unaffable" teacher, he would experience more anxiety than a comparable low n Aff student.

The expected relationships of these personality variables to course examination performance under stress were formalized in a number of hypotheses:

1. High SA students in a high stress situation will score lower than high SA students in a low stress situation; no differential response to stress will be shown by moderate or low SA students.
2. Low S students in a high stress situation will score lower than similar students in a low stress situation; no differential response to stress will be shown by moderate or high S students.
3. Moderate n Ach male students in a high stress situation will score lower than similar males in a low stress situation; no differential response to stress will be shown by low or high n Ach males. (No prediction was made for female students.)
4. High n Aff students in a high stress situation will score lower than similar students in a low stress situation; no differential response to stress will be shown by moderate or low n Aff students.

Because of the uncertainty of the validity of these relatively new personality measures, it was decided that the significance of differences between groups should be tested with a two-tailed test.

Subjects. The *S*s were students in Introductory General Psychology at the University of Michigan during the spring semester, 1955. A regression equation was devised to predict the scores of the first two course examinations. The weights for these indepedent variables in the regression equation were grade-point average, American Council on Education test total score, and scores on the first two course examinations. The weights for these independent variables were computed from data on students in the same course during the previous semester.

When the predicted third examination scores were computed, they were used to form two equated groups by systematically assigning students to a low or high stress experimental situation where they took their third examination. This "matching" procedure controlled for certain intellectual and motivational factors which influence performance. The sample consisted of 125 students in each of the experimental groups. The high stress group consisted of 43 male and 82 female students. The low stress group contained 60 males and 65 females. In assigning *S*s to the two groups, no thought was given to their sex. The resulting discrepancy between the number of males in the high and low stress groups must be attributed to the perversity of chance.

Stress situations. The operations by which the experimental situations were to be made differentially stressful were selected with an eye to their relevance to the personality measures employed. In the *high stress* group: the examination was given in a strange room; students were very closely proctored by a strange instructor; they were told there would be a guessing penalty in scoring the multiple-choice exam; two very difficult "practice" questions preceded the examination proper; and the instructor maintained a cold, abrupt, and rather severe demeanor throughout the examination. In the *low stress* group: the students took their examination in their usual classroom; their usual instructor was with them; they were told there would be no guessing penalty; no formal instruction sheet or practice questions were given to them; they used a special answer sheet which provided a space after each question for any comments they wished to make (**8**); and the instructor tried to maintain a warm, interested, relaxed manner and to give the impression that he was sincerely interested in assisting students to do well in every way he legitimately could.

As each student finished his examination he answered a question which asked whether he had "clutched" at any time during the examination. This question was included to determine whether differences in felt anxiety would accompany the stress conditions.

Results

A chi-square test showed significantly ($p < .02$) more clutching was reported by the high stress students than by low stress students. This provided some support for the assumed differential stressfulness of the two experimental situations.

SA: The SA dimension was broken into three groups representing high, mid, and low permeability. Test performance for each of these groups of students was then analyzed under the two experimental conditions. High SA (permeable) males and females both were significantly affected by stress (Tables 1 and 2), whereas mid and low SA students were not. These findings support the

TABLE 1
EXAMINATION SCORES OF SA MALES UNDER STRESS CONDITIONS

Stress	SA	N	Mean Predicted z Score	Mean Exam z Score	Relative "Gain"
Low	Lo	17	.29	.57	.28
	Mid	17	−.03	−.21	−.18
	Hi	20	−.20	.19	.39*
High	Lo	11	.10	.44	.34
	Mid	20	.04	.12	.08
	Hi	12	.13	.00	−.13*

* Significance of difference between starred values: $t = 2.26$, $p < .05$ (2 tails).

TABLE 2
EXAMINATION SCORES OF SA FEMALES UNDER STRESS CONDITIONS

Stress	SA	N	Mean Predicted z Score	Mean Exam z Score	Relative "Gain"
Low	Lo	22	.26	.21	−.05
	Mid	21	.12	.05	−.07
	Hi	18	−.17	.14	.31*
High	Lo	21	.28	.26	−.02
	Mid	27	.08	−.16	−.24
	Hi	33	−.13	−.38	−.25*

* Significance of difference between starred values: $t = 2.43$; $p < .02$ (2 tails).

hypothesis that stress is more detrimental to the examination performance of highly permeable students than it is to that of moderately or relatively impermeable students.

S: The hypothesis with regard to the stability dimension received only tenuous support from the data. Low stability males displayed a performance difference under stress in the predicted direction (Table 3). The t of 1.97, however, is not very reliable ($.10 > p > .05$; 2 tails). No difference approaching an acceptable level of significance was found in the case of females.

n Ach: The prediction with regard to need for achievement was unsupported. No differences in performance for high, mid, or low n Ach males under the two stress conditions were observed. Analysis of the females' data, however, showed a significant differential

TABLE 3

EXAMINATION SCORES OF S MALES UNDER STRESS CONDITIONS

Stress	S	*N*	Mean Predicted *z* Score	Mean Exam *z* Score	Relative "Gain"
Low	Lo	19	−.03	.37	.40*
	Mid	17	−.01	.03	.04
	Hi	19	.08	.15	.07
High	Lo	15	.13	.10	−.03*
	Mid	15	.03	.37	.34
	Hi	13	.08	.01	−.07

* Significance of difference between starred values: $t = 1.97$; $p < .10$ (2 tails).

TABLE 4

EXAMINATION SCORES OF n ACH FEMALES UNDER STRESS CONDITIONS

Stress	n Ach	*N*	Mean Predicted *z* Score	Mean Exam *z* Score	Relative "Gain"
Low	Lo	21	.07	.34	.27*
	Mid	14	−.01	.12	.13
	Hi	20	.01	.15	.14
High	Lo	21	−.04	−.54	−.50*
	Mid	24	−.10	−.21	−.11
	Hi	21	.17	.11	−.06

* Significance of difference between starred values: $t = 2.96$; $p < .01$ (2 tails).

reaction to stress by the low n Ach females (Table 4) such that their performance was detrimentally affected by stress.

n Aff: The hypothesis with regard to the n Aff dimension was not supported by the data. Neither males nor females displayed performance differences to stress related to their n Aff scores.

SA-S types: Since the SA and S measures correlated only .12 for males and .11 for females, another way of extracting meaning from them is to treat them as essentially noncorrelated dimensions defining four personality types—one at each of the four resulting quadrants (**14**). Median breaks were made on each dimension, and *S*s scoring within one point of either median were discarded. The remaining *S*s were classified as one of four SA-S types. Analysis of performance difference of these types showed no dif-

TABLE 5

EXAMINATION SCORES OF SA-S TYPE FEMALES UNDER STRESS CONDITIONS

Stress	Type			Mean Predicted z Score	Mean Exam z Score	Relative "Gain"
	SA	S	N			
Low	hi	hi	14	−.06	.25	.31*
	lo	hi	13	.31	−.09	−.40
	lo	lo	10	.12	.29	.17
	hi	lo	9	−.25	−.30	−.05
High	hi	hi	19	−.15	−.41	−.26*
	lo	hi	9	.43	.36	−.07
	lo	lo	11	.22	.25	.03
	hi	lo	19	−.18	−.49	−.31

* Significance of difference between starred values: $t = 2.04$; $p < .05$ (2 tails).

TABLE 6

EXAMINATION SCORES OF ACH-AFF TYPE MALES UNDER STRESS CONDITIONS

Stress	Type			Mean Predicted z Score	Mean Exam z Score	Relative "Gain"
	Ach	Aff	N			
Low	hi	hi	13	−.09	−.16	−.07
	hi	lo	9	.36	.44	.08
	lo	lo	17	−.07	.37	.44*
	lo	hi	15	.10	.11	.01
High	hi	hi	12	.08	.04	−.04
	hi	lo	9	.18	.22	.04
	lo	lo	10	.00	.39	.39**
	lo	hi	5	.24	.35	.11

* $t = 2.75$; $p < .02$ (2 tails).
** $t = 1.70$; $p < .15$ (2 tails).

ferential reaction to stress for males. However, female *Ss* (Table 5) who were high SA and high S types showed significantly depressed performance under stress.

Ach-Aff types: A similar type analysis was made for the n Ach and n Aff measures which correlated only .08 and .10 for males and females respectively. The males (Table 6) showed no differential reaction to stress, but an interesting observation was an apparent improvement of low n Ach-low n Aff males under both experimental conditions. The female *Ss* (Table 7) who were low n

TABLE 7
EXAMINATION SCORES OF ACH-AFF TYPE FEMALES
UNDER STRESS CONDITIONS

Stress	Type			Mean Predicted z Score	Mean Exam z Score	Relative "Gain"
	Ach	Aff	N			
Low	hi	hi	16	−.12	−.01	.11
	hi	lo	10	.26	.41	.15
	lo	lo	12	.29	.82	.53*
	lo	hi	17	−.15	−.07	.08
High	hi	hi	14	.04	−.05	−.09
	hi	lo	14	.21	.19	−.02
	lo	lo	19	.07	−.32	−.25*
	lo	hi	19	−.10	−.40	−.30

* Significance of difference between starred values: $t = 2.60$; $p < .02$ (2 tails).

Ach-low n Aff showed significantly depressed performance under stress.

Discussion and Conclusions

The belief that individual performance differences in response to stress during examinations are related to detectable personality variables has received some support from this study. One conclusion suggested by our analysis is that there may be a sex difference in response to stress. The greater frequency of significant findings for females is notable. Mead (**9**) has proposed that the role of the American female is social and dependent behavior, not career accomplishment in competition with males. Field (**2**) and McClelland et al. (**7**) have reached a similar conclusion. If we can regard many female college students as persons who have not accepted academic achievement as one of their strong intrinsic motives, it may be possible to regard them as relatively poorly equipped to cope with stress in achievement situations which they perceive as only incidentally related to the gratification of their social motives. The implication here is that, if an activity is relatively unimportant or uncentral to one's dominant motives, one is poorly stimulated to develop those skills or defenses which keep the performance of this activity relatively unaffected by stress.

The writer hesitantly raises the point that, if this latter hypothesis is true, it may mean that it is easier to produce stress effects in contrived laboratory experiments than it is in experimentally manipulated "real life" situations in which the *S*s are highly involved and to which they bring well-practiced defenses.

The SA (permeability) dimension has some promise as a means of differentiating individual performance response to stress. Since the conceptual meaning of this dimension includes sensitivity to influences from others, it is possible that the highly permeable student was most affected by the experimental manipulations which involved the behavior of the instructor during the examination. However, since the design included a number of stress manipulations, specification of the specific effects of any one is conjectural.

The finding that the high SA-high S female students showed the greatest performance decrement under stress of all the SA-S types

SA-low S ty The high

inative, dependent, disorg

The high SA-high S student has

tivity, changeability, sociability, impu

of these types are high on SA and, henc

performance decrement under stress

son is relatively stable while

tively anxious. It mi

SA-low S fema

stress th

the experi-

the low stress situa-

uch in the high stress situa-

an adequate "spread" on stress for

of this interpretation is the fact that

amination score of these students was the lowest

in both stress groups.

ch variable has shown a sex interaction. While the n

hesis with regard to males has found no support, low

les displayed a significant tendency to poorer per-

tress. It may be that the earlier proposal that

nhibitory conflict in academic

achievement situations is especially true of low n Ach females who may not have successfully resolved the conflict in favor of accepting academic achievement as one of their strong motives.

The n Aff variable has shown no independent discriminatory value in this study. The hypothesis with regard to n Aff is not supported.

The Ach-Aff type analysis revealed no relationship of these types to examination performance under stress for males. The apparent improvement, regardless of stress on this test relative to two previous tests, of the low n Ach-low n Aff males may indicate a kind of student who tends to "put off" study effort until rather late in a semester when the threat of failure induces intense motivation to study.

The low n Ach-low n Aff females did respond differentially under stress. This indicates that, although the n Aff variable has no predictive value when used alone, when combined with the n Ach measure it does help indentify a type of female adversely affected by examination stress.

[illegible]ognized that a student's performance on an exami- [illegible]mined by many factors. Intelligence, amount of [illegible]bject matter, and familiarity with test-taking [illegible]nly accepted as determining influences. Our [illegible] personality characteristics of the student and [illegible] of the testing situation must also be considered [illegible]ne performance.

Summary

The relationship of four personality variables to student performance on course examinations under experimentally manipulated stress was investigated. The findings were that the students most detrimentally affected by the stress conditions employed here were: highly permeable students, low stability males, females with low achievement motivation, females of high permeability combined with high stability, and females of low achievement motivation combined with low affiliation motivation. Also, stress had a greater detrimental effect on females than on males.

References

1. Deese, J., Lazarus, R., & Keenan, J. Anxiety, anxiety reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55–60.

The writer hesitantly raises the point that, if this latter hypothesis is true, it may mean that it is easier to produce stress effects in contrived laboratory experiments than it is in experimentally manipulated "real life" situations in which the *S*s are highly involved and to which they bring well-practiced defenses.

The SA (permeability) dimension has some promise as a means of differentiating individual performance response to stress. Since the conceptual meaning of this dimension includes sensitivity to influences from others, it is possible that the highly permeable student was most affected by the experimental manipulations which involved the behavior of the instructor during the examination. However, since the design included a number of stress manipulations, specification of the specific effects of any one is conjectural.

The finding that the high SA-high S female students showed the greatest performance decrement under stress of all the SA-S types is somewhat contrary to what might have been expected. The high SA-low S type of student has been described as "sensitive, imaginative, dependent, disorganized, uncontrolled, and cycloid" (**15**). The high SA-high S student has such characteristics as sensitivity, changeability, sociability, impulsivity, and stability. Both of these types are high on SA and, hence might be expected to show performance decrement under stress. But the high SA-high S person is relatively stable, while the high SA-low S person is relatively anxious. It might have been expected, then, that the high SA-low S female students would have performed less well under stress than the high SA-high S female students. The results indicate just the reverse. It may be that the high SA-low S female students were already so near maximal tension that the experimenter was unable to reduce their tension in the low stress situation and/or unable to increase it much in the high stress situation—and, hence, failed to get an adequate "spread" on stress for these people. In support of this interpretation is the fact that the predicted examination score of these students was the lowest of the types in both stress groups.

The n Ach variable has shown a sex interaction. While the n Ach hypothesis with regard to males has found no support, low n Ach females displayed a significant tendency to poorer performance under stress. It may be that the earlier proposal that female students experience an inhibitory conflict in academic achievement situations is especially true of low n Ach females who may not have successfully resolved the conflict in favor of accepting academic achievement as one of their strong motives.

The n Aff variable has shown no independent discriminatory value in this study. The hypothesis with regard to n Aff is not supported.

The Ach-Aff type analysis revealed no relationship of these types to examination performance under stress for males. The apparent improvement, regardless of stress on this test relative to two previous tests, of the low n Ach-low n Aff males may indicate a kind of student who tends to "put off" study effort until rather late in a semester when the threat of failure induces intense motivation to study.

The low n Ach-low n Aff females did respond differentially under stress. This indicates that, although the n Aff variable has no predictive value when used alone, when combined with the n Ach measure it does help indentify a type of female adversely affected by examination stress.

It is recognized that a student's performance on an examination is determined by many factors. Intelligence, amount of knowledge of the subject matter, and familiarity with test-taking procedures are commonly accepted as determining influences. Our results indicate that personality characteristics of the student and the stressfulness of the testing situation must also be considered to determine performance.

Summary

The relationship of four personality variables to student performance on course examinations under experimentally manipulated stress was investigated. The findings were that the students most detrimentally affected by the stress conditions employed here were: highly permeable students, low stability males, females with low achievement motivation, females of high permeability combined with high stability, and females of low achievement motivation combined with low affiliation motivation. Also, stress had a greater detrimental effect on females than on males.

References

1. Deese, J., Lazarus, R., & Keenan, J. Anxiety, anxiety reduction, and stress in learning. *J. exp. Psychol.,* 1953, **46,** 55–60.

2. Field, W. F. The effects on thematic apperception of certain experimentally aroused needs. Unpublished doctoral dissertation, Univer. of Maryland, 1951.
3. Korchin, S. J., & Basowitz, A. Perceptual adequacy in a life stress. *J. Psychol.*, 1954, **38,** 495–502.
4. Lazarus, R., Deese, J., & Hamilton, R. Anxiety and stress in learning: the role of intraserial duplication. *J. exp. Psychol.*, 1953, **47,** 111–114.
5. Lazarus, R., & Eriksen, C. W. Effects of failure stress upon skilled performance. *J. exp. Psychol.*, 1952, **43,** 100–105.
6. Lucas, J. D. The interactive effects of anxiety, failure, and intraserial duplication. *Amer. J. Psychol.*, 1952, **65,** 59–66.
7. McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
8. McKeachie, W. J., Pollie, D., & Speisman, J. Relieving anxiety in classroom examinations. *J. abnorm. soc. Psychol.*, 1955, **50,** 93–98.
9. Mead, M. *Male and female.* New York: William Morrow, 1950.
10. Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45,** 91–96.
11. Osler, S. F. Intellectual performance as a function of two types of psychological stress. *J. exp. Psychol.*, 1954, **47,** 115–121.
12. Polly, S. Cognitive rigidity as a function of threat. *J. Pers.*, 1955, **23,** 346–355.
13. Smith, D. E. P., & Raygor, A. L. Verbal satiation and personality. *J. abnorm. soc. Psychol.*, 1956, **52,** 323–326.
14. Smith, D. E. P., & Wood, R. L. Noncognitive personality dimensions as indices of reading improvement at the college level. *Amer. Psychologist,* 1953, **8,** 438. (Abstract)
15. Smith, D. E. P., Wood, R. L., Downer, J. W., & Raygor, A. L. Reading improvement as a function of student personality and teaching method. *J. educ. Psychol.*, 1956, **47,** 47–59.
16. Smock, G. D. The influence of psychological stress on the "intolerance of ambiguity." *J. abnorm. soc. Psychol.*, 1955, **50,** 177–182.
17. Taylor, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41,** 81–92.
18. Taylor, J. A., & Farber, I. E. The effect of failure and success as a function of ascendancy and submission. *Amer. Psychologist*, 1948, **3,** 361. (Abstract)
19. Taylor, J. A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44,** 61–64.

Received July 29, 1957.

PROBLEMS OF PROGENY RELATED TO PATERNAL EDUCATION

JAMES DRASGOW

University of Buffalo

During the process of studying psychological breakdowns among college students (**1**), it was noticed that they did not present the often observed and more obvious symptoms of anxiety or worry that so often precede breakdowns at the adult level. The overt expression of anxiety and worry appeared to follow rather than precede the breakdown episodes. In every instance, the student's college career was prematurely closed.

Another study (**2**) specifically investigated the premature closing of college careers; one of the findings was that the academic mortality rate was significantly less for students who lived in residence halls than it was for those who lived at home. While lasting longer and experiencing fewer breakdowns, the dorm students *expressed* more anxieties and worries on a simple "worry check list." This latter finding reinforced the earlier observation that those who broke down did not show manifest symptoms of anxiety prior to collapse. These previously unrelated results would appear to support the popular theoretical belief that people who are able to express their anxieties may be "draining off" whatever it might be in others that is supposed to stimulate a culminating collapse.

The present article is aimed at one of the more measurable aspects of the earlier researches. It had been found that not only did the residence hall students survive longer and express more worries, but they also had fathers with more education. For example, the majority of the fathers of the dormitory students had been exposed to at least one year of college, while the majority of the fathers of the students who lived at home had not succeeded in finishing high school; the difference was significant well beyond the .01 level.[1] This finding fostered the problems of crossvalidation and determination of degree of association, if any, between paternal education and number of worries professed by progeny. The present project focuses on these issues.

---

[1] No statistically significant relationships of discrimination or association were found for maternal education.

Procedure

The folders of 50 students were drawn sequentially from the university's alphabetically ordered files. Extent of paternal education was obtained from Personnel Questionnaires that entering students completed; the number of worries was obtained from the questionnaire's "worry check list" which is reproduced in Figure 1. Each student's Minnesota Multiphasic Personality Inventory (MMPI) was classified as "stable" or "unstable" according to the following criteria: stable profiles had all scales within normal

	Considerable or Frequent Anxiety	Disturb Me Moderately	Little or No Worry
1. Poor health			
2. Fear of rats, mice or snakes			
3. Unfair treatment by teachers			
4. Harmony in the home			
5. Experiencing a sense of inferiority			
6. Being quite depressed when ignored			
7. Preferring recreation alone			
8. Being "moody"			
9. "Day-dreaming" while working			
10. Occupational possibilities			
11. Having my feelings easily hurt			
12. Finances at home			
13. Difficulty in making friends			
14. Never being asked to direct things			
15. Difficulty in concentrating			
16. Often depressed and in low spirits			
17. Slow to forget unpleasant experiences			
18. No enthusiasm about life's possibilities			
19. Worrying about little things			
20. Suspecting people of "underhanded" motives			
21. Family supervision and approval of behavior			

FIG. 1. THE WORRY CHECK LIST

TABLE 1
RELATIONSHIP BETWEEN FATHERS' EDUCATION AND THEIR COLLEGE CHILDREN'S EXPRESSED WORRIES WHEN THE TWO VARIABLES ARE CUT TO MAXIMIZE DISCRIMINATION

		Fathers' Education	
		College	No College
Worries	>3	12	13
	$\leqq 3$	3	22

limits except *Mf* which is often elevated among college students, and $F - K < -12$ (**3**). All other profiles were regarded as unstable (**4**).

Results

A discriminant function analysis between the number of worries and the extent of paternal education resulted in an optimum maximizing cutoff point for discrimination as a function of worry at the three-worry level; the resulting 2×2 matrix is presented as Table 1. The cosine- pi tetrachoric correlation coefficient on this matrix is .66, and χ^2 is significant beyond .01. With these quantitative results, it may be qualitatively stated that there is indeed a sizeable and significant relationship between the number of worries expressed by a student and the number of years of education his father has had: the more educated the father is, the more worries the progeny profess. An analysis of the variance of worries among students resulted in a statistically insignificant F. This indicated that no one particular worry or set of worries listed in Figure 1 would serve to discriminate the children of parents with college from the children of parents without college.

The finding of $F > .05$ for worries among students was corroborated by the results obtained with the MMPI. The MMPIs, when classified as stable or unstable according to the previously indicated criteria, showed no statistically significant relationship to either number of worries indicated on the check list or to amount of paternal education.

Conclusion

A formulation from all the foregoing is that the child of a college parent may be more motivated for college than the child of a groups in Brigham's book is clearly contradictory to the experience of the present writers.

Method

Subjects

After the termination of the German-Polish war of 1939, a considerable part of the Polish forces, together with thousands of civilians, withdrew to the neighboring states of Hungary and Rumania. Subsequently, almost every member of this group was evacuated to France according to the prewar alternate plan formulated by the Polish government. When, later in the war, France experienced the same fate as her Polish ally, nearly all these Polish people, military and civilian, escaped to the British Isles. Following the Stalin-Sikorski Agreements of 1941, affecting over 1,000,000 Polish citizens previously deported to the Soviet Union, a second substantial group of Poles was brought to the United Kingdom. Again in 1944, after the Allied invasion of the continent, many Polish people who had been unwilling conscripts in the German army and labor battalions were reunited with the Polish national forces based in Great Britain. Finally, in 1945, after the German army had capitulated, over 100,000 members of Poland's famous General Ander's Corps, took up their station in England. These men were later joined by cilivians who had previously found refuge in the Near East. Thus in 1945, there were in Great Britain about 400,000 Poles. All classes and all parts of Poland were represented, and this group was generally regarded as a fairly representative sample of the Polish nation.

The first constructive approach to the field of applied psychology under Polish auspices was the Polish Military Psychodemographical Institute, organized in 1944, which had its central office in London, a branch office in Edinburgh, and several selection board centers distributed throughout the United Kingdom. Several officers, already possessing extensive educational backgrounds, completed a specialized training program in psychometric theory and practice at Edinburgh University under the direction of the late Godfrey H. Thomson and the late James Drever. To supplement this group of highly trained personnel, a number of psychological examiners were trained by the institute. In addition, a few of the former group continued their studies at the universities of Edinburgh and London. These, while engaged in their graduate studies, were also carrying on research projects dealing with the measurement of intelligence, apitudes, and personality factors as exhibited by the Polish population.

The Polish sample, tested by one of the writers (Wysocki) and his colleagues according to their military assignments, consisted of 15,267 men, including 651 officers. In addition, 666 adults (597 men and 69 women) were tested by the same writer in the following groups: 88 male students of a Polish university, 100 male graduates of Polish high schools, 232 male and 39 female high school seniors, 105 men who had completed elementary school education, 36 Polish elementary school teachers, 36 Signal Corps soldiers, and 30 members of the Polish Women Navy Auxiliary Service (WAVES).

Test

Raven's Progressive Matrices (**5**) were administered to all *S*s. The test has been widely used in England since 1938 and was considered to be particularly effective in this context because of its nonverbal character.[2]

Raven's test consists of 60 matrices from each of which a part is missing. The *S*s has to choose the missing part to complete a pattern from six to eight alternatives drawn below each pattern. There are five sets of 12 patterns each. The first series requires mainly ability in accuracy of discrimination, but later series become progressively more difficult requiring capacity to grasp analogies, capacity to form comparisons and other logical relations. A person's intellectual capacity is assessed on the basis of his total score. The test correlates highly with Terman-Merrill scale ($r = .86$) and has a high saturation (.82) with the Spearman's g factor. The test may be administered with or without a time limit, and it can be used either as an individual test or as a group test.

The time limit of 20 minutes, as used by the British armed forces, British industrial institutions, and British schools, was implemented in all cases from which the following data were derived.

Results and Discussion

The results obtained by the writer's colleagues are reproduced in Table 1, and a few groups tested by the present writer (Wysocki) are specified in Table 2.[3]

As one would expect, Tables 1 and 2 show significant differences in test scores between officers and enlisted men, as well as among several selected groups of *S*s. It is interesting to note that various groups of Polish nationals, as specified in Table 2, are ranked in about the same order as this test ranked similar English groups (**14**). If we look later at Tables 3 and 4, we also find (*a*) a similar hierarchical order of groups of *S*s selected according to

---

[2] Coefficient of correlation with British Army S.P. 15 (verbal intelligence tests adapted by Wysocki for Polish population) was $r = .697 \pm .036$

[3] Groups 4, 6, 10, and 11 are also included in Table 1.

TABLE 1

RAVEN'S MATRICES (1938) MEAN SCORES OF POLISH OFFICERS AND ENLISTED MEN

Ss	*N*	Mean raw score	*SD*	SD_M	Mean IQ
Officers	651	45.20	7.43	.29	122.7
Enlisted men	14,616	32.50	10.39	.09	103.0

TABLE 2

RAVEN'S MATRICES (1938) MEAN SCORES OF SEVERAL GROUPS OF POLISH ADULTS

Ss	Mean raw score	Mean standard score	Mean IQ
University students (88 men)	48.64	1.85	130
High school graduates (100 men)	45.59	1.55	124
High school seniors (39 girls)	42.78	1.28	120
Military staff school students (87 officers)	42.93	1.30	120
High school seniors (232 boys)	41.00	1.11	117
Cadet-officers (564 men)	39.58	.97	116
Elementary school graduates (105 men)	39.13	.93	115
Elementary school teachers (36 men)	35.64	.59	109
Signal corps soldiers (36 men)	33.94	.43	106
Sailors (727 men)	30.71	.12	102
Infantry n.c. officers (598 men)	28.68	−.08	99
WAVES (30 women)	26.00	−.29	95

other than military criteria and (*b*) similarity with results obtained from the intelligence testing of English speaking *Ss* in Great Britain and in this country (**5, 7, 9, 12, 13, 14**). This shows that intelligence test scores of Polish *Ss* have distributions similar to those of English speaking populations and are quite comparable not only with corresponding English scores but also with those yielded by other intelligence tests in Great Britain and the United States for similar groups.

The reader, however, may wonder at the relatively low IQ levels of the Polish elementary school teachers as shown in Table 2. This may be explained by the fact that in prewar Poland elementary school teachers were trained in the five-year teachers' training schools (*Seminarium*) after having completed a seven-year elementary school. Thus, their education was on a high school level and not, as in this country or in Great Britain, on a college level. The standard of the Polish elementary school teachers'

TABLE 3

DISTRIBUTION OF IQs OF POLISH ENLISTED MEN ACCORDING TO THE LEVELS OF EDUCATION

School	Grade completed	*N*	Percentage of sample	Raven's (1938) mean raw score	Mean Stand. score	Mean IQ
Elementary schools	1	585	4.00	26.57	−.28	95
	3	1,389	9.50	26.92	−.25	96
	4 and 5	3,537	24.20	28.31	−.11	98
	6 and 7	4,443	30.40	29.22	−.02	100
Lower technical and commercial schools	10	2,558	17.50	35.00	.53	108
Secondary schools	11	1,885	12.90	38.94	.91	115
Academic schools	12 and above	219	1.50	42.55	1.26	120

education was much lower than that of those students who graduated from the Polish secondary schools, called *Gymnasium*, preparing their students for subsequent university studies. Table 2 also shows that the IQs of the army higher rank officers are lower than the IQs of high school (*Gymnasium*) graduates. This lends support to the present writers' contention that the selection to the higher officer ranks in the Polish army was not always carried out on an objective basis.

The test results obtained from enlisted men were subjected to percentage analysis with regard to several characteristics of the Polish *S*s tested. Table 3 shows the distribution of IQ points with regard to educational levels. Here, again, one may observe a hierarchical distribution of IQs of *S*s according to levels of education, from elementary school at the bottom to universities at the top. Again, an English intelligence test (Raven's matrices) ranked the Polish *S*s in about the same way as it did English speaking *S*s (**5, 8, 9**).

Table 4 shows the distribution of IQs with regard to occupational backgrounds of Polish adults. The differences between urban and rural intelligence in this country and elsewhere were widely discussed in the last two decades (**2, 3, 6, 10**). Here, in Table 4, the marked differences in mean scores between agricultural and industrial groups seem to confirm the influence of urban and rural environment in the Polish population also. Table 4 also shows significant variations of intelligence with regard to dif-

TABLE 4

DISTRIBUTION OF IQs OF POLISH ENLISTED MEN ACCORDING TO OCCUPATIONS

Occupation	*N*	Percentage of sample	Raven's (1938) mean raw score	Mean standard score	Mean IQ
Domestic services	44	.30	26.76	−.27	95
Agriculture	9,281	63.50	27.97	−.15	97
Industry	2,952	20.20	32.51	.29	105
Transportation	234	1.60	33.06	.34	106
Public services	219	1.50	34.25	.46	108
Commerce	59	.40	38.22	.75	112
Other occupations	1,798	12.30	40.03	1.07	116
Schools	29	.20	42.03	1.21	119

TABLE 5

RAVENS MATRICES (1938) MEAN SCORES OF ENGLISH (9) AND POLISH ADULTS ACCORDING TO AGE LEVELS

*S*s	*N*	Chronological age in years					
		20	25	30	35	40	45
English[a]	5,857	44	44	42	40	38	35
Polish	3,247	31	31	29	27	25	24

[a] No time limit for English adults (9, p. 15).

ferent occupational groups—again, similar to the results obtained with the same test and the same time limit by British authors (**12, 13, 14**). Also, the age groups show marked differences in their responses to Raven's intelligence test. The decline in test scores in our sample, as shown in Table 5, may be observed already from the age of 30, and the men of 40 or 50 show a still greater decline of scores than do 30-year-old enlisted men. This characteristic decline of intelligence levels as measured by Raven's Progressive Matrices for different age groups may also be noted in Raven's test scores of British adults (**9, 12**).

The results obtained in this investigation show that scores on intelligence tests should be treated with caution. The test score alone is not necessarily a reliable index of intelligence level, particularly when one attempts to draw conclusions from the comparing of scores of different national groups. There are other factors such as education, culture, socioeconomic status, and emo-

TABLE 6

RAVEN'S MATRICES (1938) MEAN SCORES OF ENGLISH AND POLISH MEMBERS OF THE ARMED FORCES

Ss		English	Polish	*D*	SD_D	*t*
Officers	*N*	643	651			
	M	46.00	45.20	.80	.37	2.18
	SD	5.70	7.43			
	SD_M	.22	.29			
Enlisted men	*N*	15,000	14,616			
	M	34.00	32.50	1.50	.11	14.15
	SD	7.70	10.39			
	SD_M	.06	.09			

tional reactions which intelligence tests do not measure, but which may affect intelligence test scores to a marked degree. One of the present writers has received from the British War Office Selection Center (**11**) data on Raven's Progressive Matrices (1938) obtained from the testing of English officers and enlisted men. The test was administered to English adults at the same time our data from the Poles were being collected, and the same procedure of the test administration was applied. Thus, the English data were quite comparable to the data reproduced in the present investigation. Table 6 shows the critical ratio between the English and Polish officers ($t = 2.18$) statistically insignificant, and the critical ratio between the English and Polish enlisted men ($t = 14.15$) significant at the .01 level in favor of the former. Standard deviations of both groups indicate, however, that the groups of English *S*s were more uniform than the groups of Polish *S*s. In other words, the Polish samples were characterized by the wider range of intelligence levels which increased the critical ratio.

Let us now compare the English and Polish samples in Brigham's study of 411 and 383 *S*s respectively (**1**, pp. 124, 126, 142). Mean combined score for the English was 14.87 and for the Poles 10.74, and the critical ratio 33.4 in favor of the former. In our investigation, there was no significant difference between the officers and not so large a difference between enlisted men of both national groups. In the majority of cases, the Polish soldiers included in the present study encountered psychological tests for the first time in their lives, whereas the English soldiers were already familiar with testing while attending schools. Jones, discussing the same problem with regard to the American *Ss* writes: "... allowance for town familiarity with tests would give a difference of town and country of five points, which is in agreement with the eventual conclusion on the American results" (**6**, p. 5).

Fowler in his study (**4**) on 10-year-old pupils from 14 elementary schools of Detroit and Hamtramck, Michigan, has found no significant differences in mental test results of the Polish and white American nonethnic groups. Whereas in the Brooklyn, New York, area intelligence tests results of Polish children are consistently lower than those of American nonethnic children. These results point out the effect of socioeconomic status on scores of conventional intelligence tests. One has to recognize psychometric limitations of intelligence tests unless we really can develop "culture-free" tests—free from cultural learnings and unaffected by socioeconomic differences.

Summary and Conclusions

Available data concerning the level of Polish intelligence (**1**) based on a small and nonrepresentative group of Polish immigrants in the United States were felt by the writers to be inadequate. The aim of this paper was to produce recent and valid data on the intelligence of Poles. The Polish sample consisted of almost 16,000 adults. As all classes and all parts of Poland were included, this sample is the most representative and also the most recent one of adult Polish intelligence.

Raven's Progressive Matrices (1938) (**9**) were used, and the test results were subjected to percentage analysis according to different military assignments, educational levels, ages, and occupational backgrounds. It is quite a striking fact that the British-made test, when administered to various groups of Polish nationals, not only ranked various Polish groups in about the same order as this test ranked similar English groups (**5, 7, 8, 9, 11, 12, 13, 14**), but also assigned scores numerically quite comparable to corresponding English scores and indeed quite comparable to those yielded by other intelligence tests in Great Britain and the United States for comparable groups. The results reproduced in this study show clearly marked differences in intelligence levels (as measured by intelligence tests) of different groups caused by different factors such as education, occupation, rural and industrial environment, and age levels similar to what we may observe in intelligence testing of English-speaking *Ss*.

Further, the results obtained from the testing of members of the Polish armed forces were compared with the results obtained by the use of the same test on the members of English units of the British Army (**11, 14**). It appeared that the differences in mean scores between Polish and English officers were statistically negligible and the differences between English and Polish enlisted men, although statistically significant, were markedly smaller than the differences between Brigham's groups of English and Polish immigrants. This lends support to the writers' contention that, the more homogenous the groups of *Ss* we compare, the smaller are the differences in intelligence levels we measure (**4**). Until really "culture-free" intelligence tests are developed—free of education, culture, and socioeconomic status influences—any comparisons of national groups based solely on intelligence tests scores should be treated with caution.

References

1. Brigham, C. *A study of American intelligence*. Princeton: Princeton Univer. Press, 1928.
2. Burt, C. *Intelligence and fertility*. London: The Eugenic Society, 1947.
3. Cattell, R. B. *The fight of our national intelligence*. London: P. S. King, 1937.
4. Dudlow, G. H. Some recent research on the Davis-Eells Games. *Sch. & Soc.*, 1956, **84,** 146–148.
5. Iscoe, I. Distribution of matrices scores of 71 college freshmen. Unpublished report, Univer. Texas, 1955.
6. Jones, H. E. Environmental handicap in mental test performance. *Univer. Calif. Publ. Psychol.*, 1932, **5**.
7. Levine, B., & Iscoe, I. A comparison of Raven's Progressive Matrices (1938) with a short form of the Wechsler-Bellevue. *Univer. Texas Report*, 1954.
8. Raven, J. C. *Progressive matrices*. London: Lewis, 1938.
9. Raven, J. C. *Guide to using progressive matrices* (1938). London: Lewis, 1956.
10. Terman, L. M., & Merrill, M. *Measuring intelligence*. London: Harper, 1937.
11. War Office Research Wing. Personal Communication. February 2, 1948.
12. Vernon, P. E. The variations of intelligence with occupation, age and locality. *Brit. J. Psychol.*, 1947, **1,** 52–63.
13. Vernon, P. E. Occupational norms for the 20-minute Progressive Matrices Test. *Occup. Psychol.*, 1949, **23,** 1.
14. Vernon, P. E. & Parry, J. B. *Personal selection in the British forces*. London: Univer. London Press, 1944.

PREDICTING INTERESTS FROM DIFFERENT KINDS OF CUES[1]

MARVIN TAYLOR

Queens College, New York

There is little doubt that effectiveness in group relations is related to the ability of the individual to perceive accurately the interests, intentions, and beliefs of other group members. Gage and his co-workers have reported positive correlations between such variables (**11, 12, 13**). Chowdrey and Newcomb found that leaders and nonleaders in various kinds of groups significantly differ in the accuracy with which they predict relevant opinions of group members (**5**). Indeed, Steiner states that one proposition underlying the rationale of teacher education is that "the more knowledge an individual has concerning the intentions, preferences, and beliefs of other persons, the more effectively he can participate in group activities with those other persons" (**15**).

Another area of research that has had a direct effect on our knowledge of social perception has been the concern with the characteristics of good and poor judges of others (**1, 3, 8, 10**). A third area of research endeavor concerned with social perception has been the attempt to find out more about the determinants of accuracy and inaccuracy of social perception (**4, 10, 11, 17**). Gronlund points out that "in the classroom, teachers are expected to judge the knowledge, attitudes, interests, and a host of other characteristics about their pupils. Although it is easier to obtain evidence in some of these areas than in others, a major share of teachers' responses are determined in part by the accuracy of their perception . . ." (**14**). Further evidence reveals that many teachers are highly inaccurate in their perceptions of students' abilities and aspirations (**7**).

Examination of the previous research suggests that one of the variables which has not been studied in any systematic way is the effect of different kinds of cues employed in social perception. For example, when a student attends a class for the first time, the teacher may perceive his motoric pattern, style of clothing, facial expressions, cleanliness, manners, verbal comments, etc., or some combination of these cues. It is probably safe to argue that social perception is mediated by both holistic and such differentiated patterns of perception. Assuming that the preceding is true, it is probable that the accuracy of an initial social perception of an individual is influenced by the mode and content of the cues available at that time. It is also probable that increments in the accuracy of social perception are influenced both by the appearance of additional cues and their order of availability. If the type of initial cues available, the appearance of additional cues, and their order of presentation could be controlled, it would seem possible to gather some valuable data about the social perception process.

---

[1] Research project completed as part of doctoral dissertation under the supervision of Francis J. Divesta, Syracuse University.

Assuming that the variables mentioned above can be controlled, the folowing null hypotheses are testable:

1. There is no significant difference between a judge's accuracy in predicting interests of a stranger resulting from exposure to different types of introductory cues.
2. Successive increments of information about a stranger do not increase a judge's accuracy of prediction of a stranger's interests.
3. Final predictions of a stranger's interests are not different from earlier ones as a result of differences in the order of presentation of cues.

Method

Subjects. One hundred and eight teacher trainees (hereafter called judges) were randomly divided into six experimental sections. The subjects were predominately female; there were 18 in each section.

Experimental situation. Ten men known to be outside the peer group of the judges, but of approximately the same age group, were tested on the Minnesota Teacher Attitude Inventory (**6**) and the California F Scale (**2**). Of these ten, two, whose test scores were consistent but with marked differences between them (i.e., one may be called an authoritarian and the other a nonauthoritarian), were requested to perform certain activities (described below) before an 8-mm. movie camera and were tape recorded in an individual interview session with the experimenter. Neither of these men was informed of the activities they were expected

TABLE 2

F TESTS OF SIGNIFICANCE BETWEEN CUE CONDITION MEANS AT EACH PREDICTION POINT

Predictions	Stranger 1			Stranger 2		
	F	*df*	*p*	*F*	*df*	*p*
One	8.44	2/105	.01	3.47	2/105	.03
Two	4.81	2/105	.01	2.43	2/105	—
Three	1.98	2/105	—	2.41	2/105	—

means on the first predictions for Stranger 1 are significant at less than the .01 level of confidence. The mean differences relevant to Stranger 2 are significant at the .03 level of confidence. The corresponding F ratios are 8.44 and 3.47 with 2 and 105 degrees of freedom (see Table 2). Differences in mean accuracy of second predictions for Stranger 1 are also found to be significant at the .01 level of confidence. On the other hand, the mean differences in accuracy of second and third predictions for Stranger 2 and third predictions for Stranger 1 are not significantly different.

The significant Fs reported above were further analyzed by means of t tests; these results are reported in Table 3. The t tests were undertaken to provide a more rigorous analysis of the differential effects that each of the cue conditions might have had upon the accuracy of the first and second predictions. The results indicate that the judges using verbal cues for their initial predictions of the interests of both strangers are significantly more accurate than

TABLE 3

COMPARISON OF RAW ACCURACY MEANS AT FIRST AND SECOND PREDICTION POINTS

Cue Sequences	Stranger 1		Stranger 2	
	t	*p*	*t*	*p*
First Prediction				
Habitual-facial	1.17	—	2.22	.05
Habitual-verbal	7.74	.01	1.44	—
Facial-verbal	6.55	.01	3.76	.01
Second predictions				
Habitual, facial-verbal, habitual	2.03	.05		
Habitual, facial-facial, verbal	3.75	.01		
Facial, verbal-verbal, habitual	2.50	.01		

the judges utilizing the facial-hand cue condition. The corresponding *ts* are 6.66 and 3.76. Furthermore, the judges' predictions for Stranger 1 are significantly more accurate when they used verbal cues than when using habitual cues ($t = 7.74$). On the other hand, when predicting for Stranger 2, no significant differences in accuracy are found between judges using verbal or habitual cues. Table 3 further reveals that no other significant differences in accuracy are found when judges predicted a second and a third time for Stranger 2. On the other hand, when predicting for Stranger 1 a second time, the judges utilizing facial-verbal cues are significantly more accurate (at the .01 level of confidence) than the judges predicting from exposure to habitual-facial or verbal-habitual cues. The second predictions of judges using verbal-habitual cues are significantly more accurate (at the .05 level of confidence) than the predictions of judges utilizing habitual-facial cues.

By employing a 3 × 3 factorial design, it was also possible to explore the effects of increments of information on predictive accuracy from one prediction to another and the effects of the different sequences of cues on the accuracy of the third prediction. The results indicate that there are significant differences between each prediction, and there is also a significant interaction between the predictions and cue conditions. These relationships are found for both strangers. The differences in accuracy when using one set of cue sequences as opposed to one of the other cue sequences reveals contradictory results. An *F* ratio of 3.28 is significant at the .05 level of confidence when predicting for Stranger 2, whereas a nonsignificant *F* ratio is reported when predicting for Stranger 1. There is no readily apparent reason why there should be differences in cue sequences for one stranger and not the other.

The finding of a significant interaction for both strangers and an examination of the accuracy means (see Table 1) seems to suport the hypothesis that verbal cues are more instrumental in affecting the accuracy of predictions than either of the other two cue conditions.

Finally, the differences between successive predictions were further analyzed by means of *t* tests. Table 5 indicates that the differences in accuracy of prediction for Stranger 1 are significant at the .01 level of confidence between first and second and first and third predictions. The differences in accuracy between second

TABLE 4
F TESTS OF SIGNIFICANCE BETWEEN THREE PREDICTIONS AND THREE CUE SEQUENCES

Source of Variation	Stranger 1			Stranger 2		
	F	*df*	*p*	*F*	*df*	*p*
Between Predictions[a]	30.65	2/210	.001	9.63	2/210	.001
Interaction: Predictions × Cues	12.64	4/210	.001	2.93	4/210	.03
Between Cue sequences	2.02	2/105	—	3.28	2/105	.05

[a] Correlation between predictions was treated as suggested by Edwards (7).

TABLE 5
COMPARISON OF MEANS FOR EACH PREDICTION

Predictions	Stranger 1			Stranger 2		
	t^a	*df*	*p*	*t*	*df*	*p*
One-Two	4.52	107	.01	2.73	107	.01
One-Three	6.04	107	.001	3.40	107	.01
Two-Three	2.12	107	.05	.59	107	—

[a] *t* test is based on correlated means.

and third predictions are significant at the .05 level of confidence. In predicting for Stranger 2, the differences in accuracy between predictions are not significant. The means and the *t* tests indicate that increasing increments of information do not have a uniform effect upon increments in the accuracy of predictions. It may be that a law of diminishing returns is in operation.

Summary

One hundred and eight teacher trainees predicted the interests of two strangers after receiving information, in different sequences, about each person. The results obtained show that the first null hypothesis does not hold. Observing a person's habitual-expressive movements and even his facial expressions when frustrated, although frequently depended on, are inadequate criteria for personality judgments as compared with listening to him speak about his feelings and aspirations. The study weakens but does not destroy the other two hypotheses tested.

References

1. Adams, H. F. The good judge of personality. *J. abnorm. soc. Psychol.*, 1927, **22**, 172–182.
2. Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.
3. Allport, G. W. *Personality: a psychological interpretation*. New York: Holt, 1937.
4. Bruner, J., & Taguri, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 2. Cambridge: Addison & Wesley, 1954. Pp. 634–650.
5. Chowdrey, K., & Newcomb, T. M. The relative abilities of leaders and non-leaders to estimate the opinions of their own group. *J. abnorm. soc. Psychol.*, 1952, **47**, 51–57.
6. Cook, W. W., Callis, R., & Leeds, C. H. *Minnesota teacher attitude inventory*. New York: Psychological Corp., 1951.
7. Cronbach, L. *Educational psychology*. New York: Harcourt, Brace, 1954.
8. Dymond, R. F. Personality and empathy. *J. consult. Psychol.*, 1950, **14**, 343–350.
9. Edwards, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
10. Estes, S. G. Judging personality from expressive behavior. *J. abnorm. soc. Psychol.*, 1938, **33**, 217–236.
11. Gage, N. L. Judging interests from expressive behavior. *Psychol. Monogr.*, 1952, **66**, No. 18 (Whole No. 350).
12. Gage, N. L. Accuracy of social perception and effectiveness in interpersonal relationships. *J. Pers.*, 1953, **22**, 14–26.
13. Gage, N. L., & Suci, G. Social perception and teacher-pupil relationships. *J. educ. Psychol.*, 1951, **42**, 144–152.
14. Gronlund, N. E. The general ability to judge sociometric status: Elementary student teachers' sociometric perceptions of classmates and pupils. *J. educ. Psychol.*, 1956, 47, 147–158.
15. Steiner, I. D. Interpersonal behavior as influenced by accuracy of social perception. *Psychol. Rev.* 1955, **62**, 268–274.
16. Stern, G G. Studies in personality typologies: The N, R, and S syndromes. Paper read at Midwest. Psychol. Ass., Chicago, May, 1953.
17. Taft, R. The ability to judge people. *Psychol. Bull.*, 1955, **52**, 1–23.

Received October 21, 1957.

SPELLING DISABILITY AND ASYNTAXIA IN A CASE INVOLVING INJURY TO THE LANGUAGE FORMULATION AREA OF THE BRAIN[1, 2]

JACK A. HOLMES

University of California

AND WILLIAM HYMAN

VA Hospital, Long Beach

In his 1955 presidential address to the APA Division of Educational Psychology, Guy T. Buswell said: "The ultimate objective, of course, is not merely to relate [psychological] research to currently accepted educational theory, but rather, to do the kind of research [in educational psychology] that will help to formulate an educational theory that is *valid*" (**1**, p. 178).

In order that learning theory in the field of educational psychology may itself be valid, however, it must be grounded in the neuroanatomy, physiology, and biochemistry of the organism that does the learning, i.e., the human being. To this end, Eames (**2**), has stressed the importance of association pathways in language disabilities, and Holmes (**6**) has worked out the relationships between the neurophysiology of the brain and the reading process. In the present paper, the writers intend to show that when, under certain circumstances, neuroanatomical evidence is combined with appropriate psychoeducational behavior, some areas of common ground may emerge.

This being the objective, the reader is asked to consider first

---

[1] Delivered before the California Educational Research Association's Annual Spring Conference, March 1957, at Santa Barbara, California.

[2] This work was done under a VA Research Fellowship at the Long Beach Veterans Hospital while the senior author was on sabbatical leave from the University of California, Berkeley. The writers express their gratitude to the VA for the use of their facilities and permission to publish case material. Likewise, the writers are indebted to J. M. Nielsen, Senior Neurological Consultant, not only for bringing this case to their attention, but also for his helpful suggestions.

the neuroclinical evidence, then the psychoeducational evidence, and finally the integration of the two.

Case History

R. J., a 35-year old, right-handed, white, male truck driver was hit with a bottle on the *left* side of his head July 14, 1956. At the receiving hospital he showed right-sided weakness, increased deep tendon reflexes on the right side, and a right Babinski sign. Angiograms, an electroencephalogram, roentgenograms, and a spinal fluid examination were reported as not remarkable, and the patient was transferred to the Veterans Administration Hospital in Long Beach. An EEG taken at this hospital was suggestive of a left-sided extra-cerebral mass—thought to be a hematoma. Before further action was taken, however, on August 26, 1956, the patient signed himself out of the hospital against medical advice.

The patient was readmitted to the VA Hospital on January 2, 1957. In the interval since his last admission, five months before, he had been bothered by transient occipital headaches, lacrimation, and diplopia. His speech had become impaired, and he complained of numbness on the right side of his face, tinnitus and pain in his right ear.

Hospital Course

A pneumoencephalogram then showed the left lateral ventricle of the brain to be moderately dilated. The neurological picture was consistent with unilateral cerebral atrophy on the left. Trephination was done bilaterally in the temporoparietal regions. Visual examination through the "burr holes" did not reveal the expected subdural hematoma.

Shortly after these diagnostic procedures had been completed, the patient began to show a gradual clinical improvement. However, two defects persisted. He had difficulty in arranging his spoken words in proper order, and he was unable to write spontaneously.

Neuroclinical Examination

The cerebral localization of the patient's present aphasia is indicated by the following clinical examination:

1. The patient is able to recognize both animate and inanimate objects; therefore, he does not have visual agnosia for objects, and we assume areas

17 and 18 to be intact. Figure 1 depicts the various areas in or on the cortex referred to during this discussion. Since R. J. is also able to recall, by revisualization, objects which he describes in color and detail, it appears that we are justified in also assuming area 19 to be intact **(7, 8)**.

2. The patient is able to understand spoken commands; therefore, he shows no auditory-verbal agnosia, and we assume Wernicke's area (Brodmann's areas 41 & 42) to be intact **(7, 8)**.

3. The patient is able to identify objects placed in his hand even though his eyes are closed; therefore, he manifests no tactile agnosia, and we assume the hand sensory area in the post-central gyrus to be intact, as is also the superior parietal.

4. The patient is able to carry out such a sequence of acts as is demanded by the command, "Light and smoke a cigarette." This would indicate that he does not have ideational or ideokinetic apraxia, and we assume that he does not have a diffuse cortical involvement or a deep lesion in the supramarginal gyrus (area 40) on the left side. Likewise, since he is able verbally to express his thoughts, he does not have cortical motor-pattern apraxia of the mouth, and we assume Broca's area (area 44) to be intact.

5. The patient's verbal responses manifest difficulties in language formulation. When asked, "What kind of work would you like to do?" he answered, "I want to drive for a truck." When inquiring as to the possibility of working again, he said, "If can work, how long and for work?" In response to a query regarding his previous employment, he said, "I was clean man [meaning janitor], and I was working truck, too." He, therefore, has obvious trouble with syntax and it *cannot* be assumed that area 37 on the major side is intact. Brodmann's area 37 is located, as Figure 1 shows, in the postero-lateral portion of the temporal lobe and, as first pointed out by Nielsen **(7)**, is concerned with the formulation of language and with diction. This area, on the major side, coordinates word-symbol impulses from various portions of the brain and organizes them into the syntactical pattern of the coherent sentence. An injury to this language formulation area would result, then, in the patient's having difficulty with syntax, grammar, rhetoric, or style **(7, 8)**.

7. The patient is able to add and subtract mentally. He can write the numbers and solve problems on paper; therefore, he shows no acalculia, and we assume his lesion has not affected the border of the interparietal sulcus or the gyri of Gratiolet **(7, 8)**.

8. R. J., as mentioned above, experiences no difficulty in writing numbers. Likewise, he can copy, using the right hand, letters, words, sentences, and drawings: therefore, he does not have agraphia for *these*, and we can assume that Exner's writing center with its corresponding motor area for the hand are intact **(7, 8)**.

It is somewhat misleading to speak of a "center" or specific "area" for any linguistic function as we have done, for actually any such function is so complicated that it takes many parts of the brain working together to execute such a verbal act, demon-

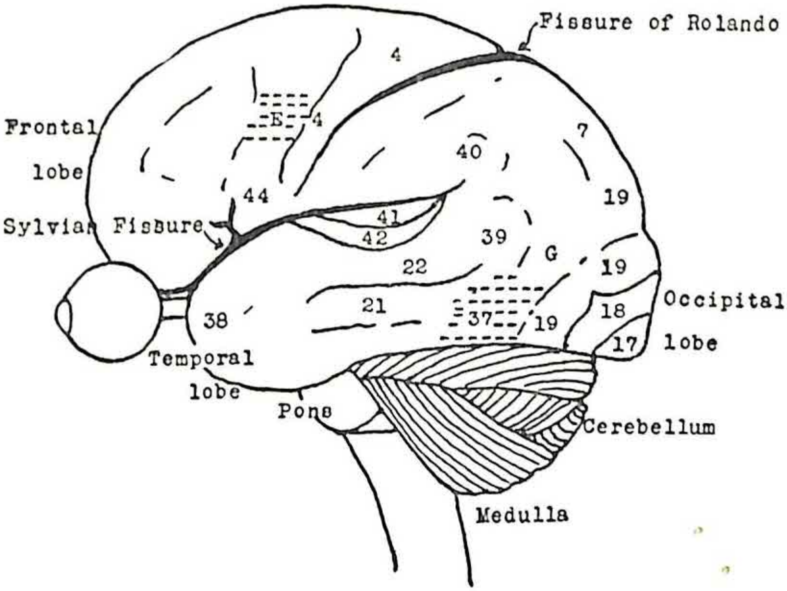

Fig. 1. Brodmann's Areas Indicated by Number (See Text for Functional Interpretation).

strate a linguistic skill, or express a symbolic thought. Such a "center" as Exner's writing center, of course, could not write by itself; on the other hand, without it, the rest of the brain could not write either.

In contrast to the patient's ability to write numbers, he experiences a great deal of difficulty in trying to write the alphabet and finds it entirely impossible to write words either singly or in a sentence. As will be shown later, this is true whether the task is initiated from an oral command, a written command, or from a self-initiated desire to express himself. From these findings, it is clear that the patient *does* have agraphia for linguistic symbols. However, since the patient can write numbers and can copy words, it is assumed that the basic difficulty rests in his inability to *formulate language* (asyntaxia), and if this is correct, it would be because he has, according to Nielsen's theory (**7, 8**), a lesion in area 37 on the major side.

Psychoeducational Examination

In order to increase our understanding of the relationship of cerebral localization to psychoeducational phenomena, the following intensive differential testing was done:

1. On a Dot-Embedded Configuration Test (3), which assesses the ability to see a dotted outline of a symbol when it is "hidden" in a background of dots, R. J. did relatively well, scoring 78 correct out of 108 items tried.[3]

2. On a Dot-Line Configuration Test (3), which assesses the ability to revisualize mentally and complete a symbol when only minimal cues are given, the patient did very poorly, scoring 30 correct out of 108 items tried. This tells us the patient has trouble with "closure" when he has to supply the "closing" information from memory.

3. On a Concept Formation Test (9), which assesses the ability to select from 10-item-sets subgroups of four words that "belong" together because they are all associated with the same concept or idea, the patient did only fairly well, scoring 29 correct, 11 wrong, and omitting 8.

4. On a Simple Language Formulation Test (9), which assesses the subject's ability to select from a choice of four possible words the best one to fill a blank left in a sentence, the patient did relatively well, scoring 43 correct out of 60 items. Note that in this test recognition rather than recall is the important requirement.

5. On a Matching Sounds Test (9), which requires the subject visually to select paired homonyms, the patient did very well in the simple part of the test, scoring 24 out of 30; but he did very poorly in the complex section, scoring only 9 out of 30 items tried.

6. In order to test the patient's limits in his ability to formulate language, a Syntax Test (9) was administered in three different forms:

Form I of the Syntax Test presents a scrambled sentence as a visual stimulus and requires the subject to select, from a visual presentation of four possible arrangements of the words, the best reformulated sentence. On this form, R. J. did fairly well, scoring 19 correct out of 25.

Form II of the Syntax Test presents the same scrambled sentences but requires the subject to reformulate the word order in his own mind, remember the order, and write the best sentences as the response for each item. The patient did very poorly on this form, scoring only 7 correct out of 25 items tried. By the use of a fixed hand mirror, the patient's eye movements were observed to make many excursions from stimulus to response during the writing of each letter. In other words, not only the memory of the word order of the sentence, but also the memory for the letter order of the word did not appear to be functioning for the purpose of writing; therefore, one is forced to question this patient's ability to spell.

Form III of the Syntax Test presents the same scrambled sentence as auditory stimuli but now requires the subject to reformulate the words in his own mind, remember the word order, and say the best sentence as his response for each item. Writing is in no way involved. On this form, R. J. was not able to give a single correct response. This was not because he could not remember the words. He could, with ease, recall and repeat the sentences in their original mixed-up order, but he simply could not straighten them out!

---

[3] The standardized norms are not applicable here, because it was felt that more information could be gained by using the instruments as power rather than as speed tests. The patient was cooperative in all tests.

From the three forms of the Syntax Test, it becomes quite evident that the real trouble is in language formulation and perhaps letter formulation or *spelling from recall*, and the difficulty is not with the writing per se.

7. To this end, then, a High School Spelling Test (**3**), consisting of 35 words of increasing difficulty, was administered to the patient in four different forms:

Form I of the Spelling Test simply required the subject to recognize and encircle the one word in four that was correctly spelled. On this mutiple-choice version, the patient did surprisingly well, scoring 20 correct responses out of 35 items. Table 1 presents, in the first four columns, the choices for the five easiest items and the patient's responses to them. It should be noted that they are all correctly encircled.

Form II of the Spelling Test required the subject to write these same words from dictation. Column 5, Table 1 shows the patient's best efforts on the first five easiest items. His total score was "0."

Form III of the Spelling Test required the subject to "peck out" the letters on the typewriter in order to spell these same words from dictation. It seemed reasonable to believe that if the patient had the letters before him he might be able to select them in the proper order to fit the formulation of at least those 20 words which Form I indicated he knew how to spell correctly. However, even this proved to be an impossible task for him—his score was "0." Column 6, Table 1 gives the responses made by R. J. on the same five items.

Form IV of the Spelling Test required the subject to spell the words aloud from dictation. In spite of the fact that writing was in no way involved, R. J. was not able to spell a single word correctly. Column 7, Table 1 gives the verbal responses made by the patient to these same five words.

8. The patient did very poorly on a Phonetic Association Test (**3**), scoring 10 out of 50 items attempted.

Discussion

There is a case reported in 1953 which in many ways is strikingly parallel to the one being reported here. In other ways, however, there are marked differences.

Trowbridge (**10**) opens his article with the interesting statement: "The syndrome of loss of ability to write with preserved ability to read, is almost unknown in the field of aphasic disorders. It is the reverse of pure word blindness (inability of the patient to read what he himself has written)."

Trowbridge's patient was a 27-year old, white, navy man who had suffered a cerebral thrombosis involving branches of the left posterior cerebral artery. Six months later, the only aphasic residuals "persisting, were disorder in spelling and writing." There was no anomia, auditory-verbal, nor visual-verbal agnosia; there was no alexia, acalculia, finger agnosia, confusion of laterality, or

TABLE 1

R. J.'S RESPONSES ON FIVE EASIEST ITEMS OF THE HIGH SCHOOL SPELLING TEST UNDER FOUR DIFFERENT CONDITIONS

Recognition from visual inspection (Multiple-choice Spelling Test)				Recall from dictation		
				Writing	Typing	Oral
1	2	3	4	5	6	7[a]
1. brav	[✓ brave]	brve	brev	barr	bcrk	da
2. dinner	dinr	[✓ diner]	denr	dinnn	diinir	dinnird
3. [✓ honey]	hony	huny	honny	hun	hucy	hogny
4. pont	pount	poynt	[✓ point]	nun	titc	poyn
5. theaf	theif	[✓ thief]	thefe	doud	eebc	thef

[a] Typed by examiner from patient's oral responses.

confusion of body scheme. He was able spontaneously to formulate speech conveying very complex thoughts. Unlike the present patient, Trowbridge's patient had no difficulty with either language or word formulation. However, both patients are similar in that, while they can copy words, neither can write them.

Trowbridge's patient could spell "automatically," but not when he "directed his thoughts toward spelling complex words." In this patient, the neurological examination indicated that the lesion was in the angular gyrus (area 39) on the major side. An unusual situation developed when the patient was asked to try to write with his *left* hand. While he insisted it was impossible, to everyone's surprise he did so very well! It was revealed, upon questioning, that the patient as a child *had been* left-handed. He learned to read before entering school. In school he was forcibly changed to right-handedness for writing. All his life he used the left hand for handling tools, but he had always written with his right hand. A left angular gyrus lesion then "produced an isolated loss of writing ability in the right hand, while the corresponding area on the right continued to perform normally for reading and other functions governed by this area. Writing with the left hand was possible as a result of the normal amount of skill present in this hand, together with the high level of functional development present in what would have been ordinarily the dominant angular gyrus" (**10**). R. J., on the other hand, was not able to write words with either hand.

Summary

A case report is presented in which the clinical evidence indicates that the patient had originally suffered a traumatic intracerebral hemorrhage, which had absorbed, with resulting impairment of area 37 on the major side. The relationships of area 37 to the other pertinent areas of the cortex concerned with aphasia are discussed and the possibility of their involvement eliminated or indicated. The psychoeducational significance, resulting from damage to the language formulation area (Brodmann's area 37) on the major side, is revealed by intensive differential testing.

1. Psychoeducationally, the most interesting finding is the evidence which indicates that the writing from *dictation type* of spelling ability and the *proofreader's type* of spelling ability are not one and the same ability. The significance of this finding takes on added weight when it is recalled that Holmes (**5**), in 1954, working with large samples of high school and college students was able to present statistical evidence indicating that these two types of spelling ability, while highly related, were in fact not the same. By a substrata-factor analysis, utilizing some of the same tests used on this case, Holmes was able to show that the subskills underlying the writing from dictation type of spelling ability depended more upon auditory images, while the proofreader's type of spelling ability depended more upon visual images.

2. One hesitates to draw major conclusions from a single case. However, the evidence in the present study seems to indicate that spelling ability as revealed by the ability *orally* to *spell* or to *write* words correctly from dictation calls upon a language formulation area (presumably area 37), while the proofreader's type of spelling ability to detect whether or not a word is spelled correctly by *visual* inspection does not require the use of this area. This being the case, then:

3. The most significant theoretical implication that can be drawn from this case study and the substrata-factor analysis of spelling ability is that, in the area of *spelling*, an empirical relationship has been established between the substrata factor theory (**4, 5**) and cerebral localization (**7, 8**).

4. The evidence indicates that agraphia (amnesic agraphia) can be caused by a basic inability to formulate the letters of words, even though the patient actually knows when they are spelled correctly.

5. The term *asyntaxia* is suggested as a word describing the linguistic loss resulting from injury to the language formulation area (area 37) of the brain.

References

1. Buswell, G. T. Educational theory and the psychology of learning. *J. educ. Psychol.*, 1956, **47**, 175–184.
2. Eames, T. H. Association pathways in language disabilities. *J. educ. Psychol.*, 1956, **47**, 8–10.
3. Holmes, J. A. *Diagnostic examination of visual perception and linguistic abilities.* Berkeley: Calif. Book Co., 1953.
4. Holmes, J. A. *The substrata-factor theory of reading.* Berkeley: Calif. Book Co., 1953.
5. Holmes, J. A. A substrata analysis of spelling ability for elements of auditory images. *J. exp. Educ.*, 1954, **22**, 329–350.
6. Holmes, J. A. The brain and the reading process. *Claremont Coll. Reading Conf. Yearb.*, 1957, in press.
7. Nielsen, J. M. *Agnosia, apraxia, aphasia: their value in cerebral localization.* New York: Paul B. Hoeber, 1946.
8. Nielsen, J. M. Agnosias, apraxias, speech, and aphasia. In A. B. Baker (Ed.), *Clinical neurology.* Vol. 1. New York: Paul B. Hoeber, 1955. Pp. 352–378.
9. Singer, H. *Linguistic tests: a battery for elementary schools.* Berkeley: Univer. Calif., 1956. (Mimeo.)
10. Trowbridge, W. V. Unilateral agraphia without alexia. *Bull. Los Angeles neurol. Soc.*, 1953, **18**, 153–155.

Received October 21, 1957.

Book Reviews

Robert W. White. *Abnormal Personality.* New York: Ronald Press. 1956. Pp. 644.

How to present the rich materials of abnormal psychology in a humanized manner, intelligible to all who can read and are interested, is the challenge the author set up for himself in the first edition published in 1948 and has again adopted as his task in the second edition published in 1956. His aim, he says, is to present the subject in such a way as to emphasize the usefulness to all students of human nature, and it is because of this desire that he sees fit to include two introductory chapters, one historical and one clinical. The chief difference in the two editions is the reduction in content devoted to consideration of dreams and hypnotism and expansion of material concerned with more recent experimental studies and what their findings implicate. The author presents similar material in a more systematic manner with increased attention to developmental stages and an attempt to take an explicitly neo-Freudian position, balancing his earlier reliance on Horney with ideas drawn from Fromm and Sullivan. The case studies used include a maladjusted adolescent person, a neurotic, a delinquent, a psychotic, and a person with a progressive brain disease; and this material is used by him to advantage in a consideration of development and integration of personality. How can we find out how the difficulties of life can make one person maladjusted, another neurotic, another delinquent, another psychotic and how another passes through comparable difficulties without troubles? To bring order, system, and understanding into facts as diverse as the illustrative cases is the purpose of the whole book.

The author shows a familiarity with the literature of all topics covered and takes an eclectic attitude toward all conflicting issues. He considers the problem of therapy fairly thoroughly for an elementary text, and sees psychotherapy not as a hopelessly complex theme, but one that has to do with removing blocks in the learning process and promoting new growth where the processes are partly understood and wholly understandable. Therapies he pays a good deal of attention to are client centered, psychoanalysis, and briefer psychoanalysis. Of nondirective therapy, he says it works on a client in just one way and neglects the possibility of using other approaches in a wise and timely fashion. He gives it as his impression that for persons with a serious neurosis, measures most patiently worked out in Freudian psychoanalysis are likely to be more promising in the long run. The basic processes considered by him as essential in all psychotherapy include (*a*) initial therapeutic relationship, (*b*) expression of feelings, (*c*) pointing out of feelings, (*d*) transference, and (*e*) behavior resulting from treatment.

There are chapters on the various mental disorders and chapters on development which are well done. The last chapter deals with the problem for society and represents a compact consideration of the meaning and implications for mankind and society of mental disease and health problems. His general conclusion is that the preventing of abnormal behavior is one in which every citizen can participate.

All in all, Dr. White has set up for himself a relatively difficult task and to a large extent has fulfilled it. To the reviewer, it is like a modernized version of Morgan's earlier efforts when less was known as given by him in *Psychology of Abnormal People*—the emphasis is on people and the knowledge given is definitely for use in understanding people. The title *Abnormal Personality* is somewhat of a misnomer since it is concerned more with people with an emphasis on diversity, rather than the thesis that there is such a thing as the abnormal personality. But that is for caption purposes only. A more honest title might be *Understanding Abnormal Personalities.*

H. MELTZER

Human Relations Research Foundation, St. Louis, Missouri

J. P. GUILFORD. *Psychometric Methods* (2nd. Ed.). New York: McGraw-Hill, 1954. Pp. 597.

When the first edition of this book appeared in 1936, specialists in educational and psychological measurement were just beginning to court mathematics and theory seriously. Thurstone headed the newly formed Psychometric Society, whose journal, *Psychometrika*, began that year. Kuder's kindred journal, *Educational and Psychological Measurement*, did not start until five years later.

Guilford's comprehensive new edition reflects changes during the eighteen-year interval and contains a great deal more material than its predecessor. New chapters on the theory of psychological measurement and psychophysical theory build a logical basis for psychophysical and psychological scaling methods. Early in the book Guilford devotes 43 pages to explaining simple mathematics and curve fitting. He warns, however, that "it was necessary to eliminate most of the statistical treatments given in the first edition." One must go elsewhere for these, he says, preferably to his *Fundamental Statistics in Psychology and Education*.

In this second edition Guilford attempts to account for a great many phenomena of human judgment by means of Helson's concept of adaptation level. The new chapters on theories of psychological tests, reliability and validity of measures, and test development serve as vehicles for presenting recent advances. Also, much updating occurs in the chapters whose titles were altered relatively little. Finally, appendixes now include tables showing significant values of r, R, t, and F and incorporate other modifications.

Among the newcomers to Guilford's pages are Cronbach, Gulliksen, and S. S. Stevens, who have become prominent since 1936. We find little reference, however, to contemporary mathematical statisticians (none to Rao, for example). Latent structure analysis seems not to be mentioned. Coombs's recent theory and techniques get only slight notice. These and similar points are consonant with Guilford's explicit disavowal of ultra-modernity: ". . . this volume represents the general field of psychological measurement as of about 1950" (p. v). But one must stop somewhere!

The procedures that Guilford explains lean rather heavily upon large-sample theory, normality, and Pearson product-moment correlation, mirroring certain limitations of statistical knowledge during the early days of psychology when psychophysical and psychological scaling methods were being developed. Professional workers and graduate students in education and psychology are indebted to Guilford for the time and effort he spent revising this textbook-handbook-manual. They will find the chapters on pair(ed) comparisons, rating scales, and factor analysis particularly valuable[1].

JULIAN C. STANLEY

University of Wisconsin

---

[1] Prospective users should read the excellent reviews by John Keats, Clemens Johnson, and Bert Green, respectively, in *Educ. psychol. Measmt.*

ERNEST R. HILGARD. *Theories of Learning:* (2nd Ed.). New York, Appleton-Century-Crofts, 1956. Pp. 563

This edition represents a thorough revision of the widely read 1948 book. Four new chapters have been added dealing with (*a*) Freud's contributions to the psychology of learning, (*b*) the use of mathematical models, (*c*) reinforcement and drive, and (*d*) "some persisting issues." These new chapters account largely for the approximately 140 added pages, although the 1948 chapter on Wheeler's contribution has been dropped. The book carries 60 pages of bibliography, and almost half of the titles did not appear in the 1948 edition.

Hilgard's general orientation, as he developed his chapters, is much the same as it was in the earlier book. His exposition is built "around the psychologists who gave currency to the several theories"—Thorndike, Guthrie, Skinner, Hull, Tolman, Wertheimer, Koffka, Köhler, Lewin, Freud, Carr and Woodworth. Most of the chapters begin with a brief and somewhat contemporaneous history of the development of the "position," which is followed by an analysis of the major experiments, a critique of the essence of the point of view, and an estimate of its bearing upon questions such as: What are the limits of learning? What is the role of practice in learning? How important are drives and incentives, rewards and punishments? What is the place of understanding and insight? Does learning one thing help you learn something else? What happens when we remember and when we forget? Each chapter is concluded with an excellent bibliography.

Hilgard writes lucidly and his approach is generally objective, thorough, and critical. He judges Hull's work to be the best effort at "hypothetico-deductive system-making" to appear in the first half of the century (p. 121). Some cogent cautions against too much emphasis upon complicated models are suggested, and the likelihood that "a good deal of rich material from psychoanalysis has eluded the academic psychologists" (p. 307) is both lamented and explained.

Theories of Learning is written for an American audience and is probably best adapted to graduate students who are taking

---

(1955, **15**, 215–218), *J. Amer. stat. Assn* (1956, **51**, 413–414), and *Psychometrika* (1955, **20**, 163–165).

their first course devoted exclusively to the psychology of learning. Hilgard's personal position is eclectic and he makes a successful effort to indicate that there is something to be gained from each and all of the theoretical positions he develops. None of the systems, in his judgment, is invulnerable to criticism and the "construction of a fully satisfactory theory of learning is likely to remain for a long time an uncompleted task."

For educational psychologists who are primarily interested in finding out more about learning that can be applied in the teaching-learning situation, Hilgard's book does not have a great deal to offer. This is true even though he concludes with a discussion of "the relation between learning and the practical demands of instruction and training" (pp. 485–90). The contention is made that there are a number of important propositions about learning on which a majority of theorists would agree, and which do have clear though rather general implications for teachers. These propositions are not intended to be revolutionary. Some of them, paraphrased a bit, go like this: 1. In deciding who should learn what, the capacities of a learner are very important. 2. Motivation that is too intense may be accompanied by distracting emotional states. 3. Learning under the control of reward is usually preferable to learning under the control of punishment. 4. Learning under intrinsic motivation is preferable to learning under extrinsic motivation. 5. Tolerance for failure is best taught by providing a background of success that compensates for experienced failure. 6. There is no substitute for repeated practice in the overlearning of skills. 7. Transfer to new tasks will be better if in learning the learner can discover relationships for himself (pp. 486f).

In connection with his discussion of practical implications, Hilgard is, in this reviewer's judgment, unusually fair and moderate in his conception of the so-called conflict between pure and applied research. He neither believes that applicational research can wait until there is a pure science, nor that the existence of the pure science will automatically result in applications (p. 489). "When the relationships between pure and applied research are properly worked out, the requisite skills are of a high order. However, the labor is divided, and there is no reason why prestige should not be more equitably distributed" (p. 490). This, from so distinguished a source, gives some support to those who are disposed to believe that one of the best ways for the teacher to learn about learning is for him to experiment with his own teaching.

STEPHEN M. COREY

Teachers College, Columbia University

PAUL K. CROSSER, *The Nihilism of John Dewey*. New York: Philosophical Library, 1955.

The most interesting, as well as the most comprehensible, part of this book will be found in the first nine paragraphs of the preface. In this passage the author presents a bill of particulars with respect to the outrages which John Dewey has perpetrated on the civilization of the West.

It is made manifest that an obsession which pursued Dewey throughout his life was nothing less than the destruction of philosophy. How near he came to achieving this insane ambition can be judged by the condition in which Mr. Crosser now finds us. The American intellectual is adrift, the American artist's outlook is fragmentary, and the American teacher is "enmeshed in unrelated details." And if this were not enough, he reports also that America's perspective has been destroyed and her influence in the world diminished accordingly.

It would be a dull fellow indeed who could suppress his curiosity over how such havoc was wrought single-handed by as mild and amiable a soul as John Dewey is reputed to have been. It is likely, therefore, that many readers will turn with anticipation from the preface to the body of the text. Those who are not in the habit of reading prefaces will be in on the secret from the beginning—that this is a monument to dullness. At least the preface reader can enjoy for a few moments his illusion that at long last the anatomy of destruction is to be laid bare.

The body of the work—232 pages of it—is devoted to shredding large chunks of three of Dewey's major works: *Logic, The Theory of Inquiry*, *Art as Experience*, and *Experience and Education*. The technique of what its author optimistically calls "this critical analysis" is to cite a sentence or phrase from Dewey and then envelop it in the heaviest sarcasm to be found outside a sophomore bull session. Sarcasm, particularly when it is prolonged for 232 pages, is difficult for even a good writer to manage—and in truth Mr. Crosser is not to be numbered among these. Sarcasm, used sparingly and with finesse, can be deadly to an opponent. But here it serves only to obscure the argument and to create an atmosphere of petty contentiousness, which would be distasteful even to one who was sympathetic to the argument itself.

The analysis is directed towards certain aspects of Dewey's theory of knowledge, his theory of art, and his theory of education. The author maintains that in all of these Dewey's nihilism is apparent. The chapter titles are indicative of the tenor of the exposition: "The Degradation of Knowledge," "The Dissipation of the Sense of Beauty," and "The Exclusion of Schools."

The blunt instrument which was used in the initial crime against the West was the kind of "cognitive indeterminableness" which is alleged to be the root of Dewey's villainy. Through it the natural and social sciences have been struck down. Subject and object, cause and effect disappear into a cognitive limbo. The inevitable result is *extreme relativism*. (Crosser always puts this term in italics lest we forget that it is a *bad thing*.)

The rest of catastrophe follows hard on this. Objects of art are consigned to outer darkness, followed quickly by the artist. The sense of beauty is lost and the very definition of esthetics becomes obscured in a miasma of *extreme relativism*.

With respect to the state of formal education the situation is, if anything, worse. The price we have paid for our cognitive undeterminableness is that the school as an institution is fast disappearing. Our concept of education as formal learning of organized knowledge is all but forgotten. In our *extreme relativism* we are even unable to discern what is educative experience and what is not. The decline of the West is all but accomplished.

It is too early to discern how telling this blow will be against the philosophy of nihilism, but it seems likely that Dewey's reputation has already been diminished in one respect. It has long been advocated in some quarters that Dewey had the worst prose style to be found in five generations. If this was true (and there are many who deny it), it is hardly true now. Dewey has met his match, and in support of this thesis the following passage from page 119 is offered in evidence:

"The cognitive violence which the removal of art objects as art objects as an analytical factor from the kind of art theory he is about to evolve is of such serious nature that the return of the art object into the fold of inquiry, after it had been left out of consideration in the evolvement of the basic principles of that inquiry, cannot restore to that kind of inquiry the character of an inquiry into the arts."

G. MAX WINGO

University of Michigan

WILLIAM E. HENRY, *The Analysis of Fantasy*, New York: John Wiley & Sons, Inc., London: Chapman & Hall, Limited. 1956, Pp. 305.

The Analysis of Fantasy is a consideration of the use of thematic apperception technique in the analysis of personality. The title is smarter as is, and to some extent can be said to be justified since the thematic apperception is one of the relatively effective methods of getting at fantasy for the sake of interpreting symbolic behavior. The material is presented in three parts. It is concerned with, respectively, the process of interpretation, illustrative analyses, and description of the pictures. There is an appendix which includes six pages of notes on the administration of the test and a very comprehensive and specifically classified bibliography of the thematic apperception technique, which takes up nearly 30 pages. Though well written, it does assume interest and persistence on the part of the reader as well as some familiarity with the background and use of the thematic apperception tests for the study of personality as a whole as well as fantasy. It is a fact, as the author says, that the interpretive significance of the themas perceived and stories told are derived from the understanding of the ways in which the emotional preoccupations are symbolized in words and plots, from the ways in which emotions bow to propriety in overt expression, or ways in which each individual lessens his own burden by expressing his own feelings through disguising them for himself and others. To the area of personality where fantasy rules, logic and propriety do not apply, and it is through the analysis of a manner in which the person manipulates the symbols of conventional communication that one derives insight into personal motives. How one learns or gains knowledge about skills directly and feelings indirectly in the use of the test is considered at some length.

Variables of form and content are treated in separate chapters.

Content refers to what the subject said and form to how he said it. And both categories are set in a conceptual framework for individual case analysis wherein comparisons are made with Rorschach formulations. The cases presented include a record of a healthy, middle class boy, a 29-year-old man, a Dakota Sioux of 14, and a Navaho girl of 11. In the chapter on the stimuli properties of the pictures, each card is described under six captions: Murray's Description; Manifest Stimulus Demand; Form Demand; Latent Stimulus Demand; Frequent Plots and Significant Variations.

This is a highly worthwhile volume for all people interested in the use of projective technique in the study of personality and for the thematic apperception tests so used particularly. Investigators of any phase of problems related to the validation or the use of the thematic apperception technique will find the bibliography very helpful.

H. MELTZER

Human Relations Research Foundation, St. Louis, Missouri

OTTO POLLAK. *Integrating Sociological and Psychoanalytic Concepts: An Exploration in Child Psychotherapy*, New York: Russell Sage Foundation, 1956. Pp. 284.

Effective application of social science concepts and approaches can make a significant contribution towards the improvement of diagnosis and therapy in child guidance work. This is the author's thesis, and the material used and the interpretations made are all slanted in the direction of proving this thesis. To a fair extent, the whole story is the report of Dr. Pollak's experience with a child guidance team. A clinical psychologist, because of illness, was not included in the team.

In this volume, the author raises a number of interesting problems that are relevant in considering ways of improving child guidance efforts and results. The much to-do about the neglect of the father in child guidance work is not new but is relevant. The concepts he advocates the use of, which in the author's way of thinking represents a reformulation of method, are four in number: family orientation, social interaction, social role, and culture conflict.

The content of the volume is presented in four parts. Part I, which is called "Family Pathology, Family Diagnosis and Family

Therapy with Child Welfare Focus," is illustrated with the presentation of two case studies. Part II, called "The Therapeutic Management of Extrafamilial Factors" is illustrated with three case studies. Part III, called "A Research Outlook" covers psychiatric literature and differences between theory and practice. Part IV is called "The Agency Looks at the Project" and is an evaluative comment by Herschel Alt, Executive Director of Jewish Board of Guardians, which, with the Russell Sage Foundation, made possible this project.

Some of the concepts stressed by Pollak are worthy of consideration. Some are over-generalizations. For example, when he says that a child welfare focus would result in consideration of situations rather than individuals so that we would not have sick husbands or sick wives, but sick marriages; not sick children or sick mothers or sick fathers, but sick families; not sick teachers or sick students, but sick school situations; not immature, sick guardians or sick administrators, but sick institutions, he is calling attention to the advisability of emphasizing emotional climate of groups rather than emotional climate of individuals; but how much reality is he running away from, as well as getting into, by so doing? It may be a fact that the "bad child" so-called, referred to by the mother as such, is best conceived as being in part made that way by the mother and all things in the form of human surroundings at home. It doesn't follow, however, that in all instances there are no such things as sick individuals if some institutions do permit, almost invite, sickness in the behavior of the people who live in them. It also doesn't follow that even in such instances that some people who come there have been made sick before they came there. The universal implication of the author's emphasis would be to have an institution to house institutions rather than individuals. To use a word which appeared in the "Talk of the Town" in a recent issue of the *New Yorker*, the author has done too much "jumboizing" of his concepts and by so doing has run off the track of reality, as well as made vivid some realities deserving attention.

On the assumption that things well done should be given survival value rather than be reformulated or thrown out, there are surprising omissions here. Included are: No mention of the work of Healy or Fritz Redl; no attention to the organization of abilities and attitudes towards the use of them; no consideration of economic factors. A greater awareness of the work realities of the clinical team and a lesser emphasis on verbal reformulations would more likely yield a more relevant and enriched exploration in child psychotherapy.

H. Meltzer
Human Relation Research Foundation, St. Louis, Missouri

Frank A. Logan, David L. Olmsted, Burton R. Rosner, Richard D. Schwartz, & Carl M. Stevens. *Behavior Theory and Social Science*, New Haven: Yale University Press, 1955.

A group of Ford fellows in behavior science with a three-year appointment at the Institute of Human Relations of Yale University have produced this volume which is an attempt to restate the Hullian point of view with regard to behavior theory and to make applications of this theory to problems in social science. The five authors of this volume, according to their own statement, met as a group five days a week throughout the first year for several hours a day and less frequently during the second and third years in which the report was written. They also acknowledge their indebtedness to a large number of other colleagues at Yale with whom they could discuss Hull's point of view and its applications.

It is indeed significant that this is one of a number of books to appear in the past two years to present an extended analysis of Hull's point of view. There has been a kind of race to see which of the theories propounded by Guthrie, Hull, Skinner, and Tolman would have the pre-eminence. It is not certain by any means that the issue has been finally settled and history will probably record that each of these psychologists has made his enduring contribution. Certainly at the present time, however, Hull's theoretical structure intrigues the largest number of psychologists and there has been greater elaboration of his theories by other psychologists than of any of the other current learning and behavior theories.

One chapter in the book is devoted to discussing the properties of the response, and another chapter is given to discussion of the stimulus and other antecedent conditions. In chapter 5, entitled "Language," the writers go beyond the basic principles underlying Hull's point of view to a discussion of thinking and problem solving. Chapter 6, entitled "Free Behavior," deals with the same phenomena that Skinner discusses under the title of operant behavior; that is, behavior which is in response to stimuli not immediately present. These authors prefer to call this free behavior, and they make a strenuous effort to fit it into the Hullian framework. Since the results are somewhat speculative, they have drawn up three models of such behavior and then state that these should be the subject for much experimentation in order to determine which of the three models comes closest to fitting actual experience. The last two chapters in the book deal with social interaction and cultural behavior. Here there is still more speculation, but always attempting to derive an explanation of social behavior from the basic Hullian proposition.

This reviewer can recommend this book to serious students of learning and behavior theory as a discussion which moves our theoretical formulation a step ahead.

PERCIVAL M. SYMONDS

Teachers College, Columbia University

PUBLICATIONS RECEIVED

GUY L. BOND AND MILES A. TINKER. *Reading Difficulties: Their Diagnosis and Correction.* New York: Appleton-Century-Crofts, Inc., 1957, pp. 486. $5.25.

WILLIAM F. BRUCE AND A. JOHN HOLDEN, JR. *The Teacher's Personal Development* New York: Henry Holt & Co., 1957, pp. 346. $4.25.

MARGARET DOREN. *Doren Diagnostic Reading Test of Word Recognition Skills.* Minneapolis: Educational Publishers, Inc., 1956, pp. 24 (paper).

ELAINE CUMMING AND JOHN CUMMING. *Closed Ranks—An Experiment in Mental Health Education.* Cambridge: Harvard University Press, 1957, pp. 192. $3.50.

LUCIEN DINTZER. *Le Jeu D'Adolescence.* Paris: Presses Universitaires de France, 1956, pp. 123 (paper).

WALTER CROSBY EELLS, Editor. *College Teachers and College Teaching: An Annotated Bibliography on College and University Faculty Members and Instructional Methods.* Atlanta, Georgia: Southern Regional Education Board, 1957, pp. 282 (paper).

JOHN EISENSON AND MARDEL OGILVIE. *Speech Correction in the Schools.* New York: The Macmillan Co., 1957, pp. 294. $4.25.

CLAUDE FRANÇOIS-UNGER. *L'Adolescent Inadapté: Réadaptation Sociale et Formation Professionnelle.* Paris: Presses Universitaires de France, 1957, pp. 302 (paper).

SOL L. GARFIELD. *Introductory Clinical Psychology.* New York: The Macmillan Co., 1957, pp. 469. $6.00.

DAVID GREENWOOD. *Truth and Meaning.* New York: Philosophical Library, 1957, pp. 114. $3.75.

STEN HENRYSSON. *Applicability of Factor Analysis in the Behavioral Sciences: A Methodolical Study.* Stockholm: Almquist & Wiksell, 1957, pp. 156 (paper).

DALE B. HARRIS, Editor. *The Concept of Development: An Issue in the Study of Human Behavior.* Minneapolis: University of Minnesota Press, 1957, pp. 287. $4.75.

ROBERT L. KAHN AND CHARLES F. CANNELL. *The Dynamics of Interviewing.* New York: John Wiley & Sons, 1957, pp. 368. $7.75.

FREDERICK KILANDER. *Health for Modern Living.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1957, pp. 494. $4.95.

GRACE LANGDON AND IRVING W. STOUT. *Helping Parents Understand Their Child's School: A Handbook for Teachers.* Englewood Cliffs, New Jersey: Prentice-Hall, 1957, pp. 508.

EDNA S. LEVINE. *Youth in a Soundless World: A Search for Personality.* New York: New York University Press, 1956, pp. 217. $5.00.

MARIAN EAST MADIGAN. *Psychology: Principles and Applications.* Second Edition. St. Louis: C. V. Mosby Co., 1957, pp. 391. $4.75.

D. J. O'CONNOR. *An Introduction to the Philosophy of Education.* New York: Philosophical Library, 1957, pp. 148. $3.75.

C. W. ODELL. *A First Course in Educational Statistics.* Dubuque, Iowa: Wm. C. Brown Co., 1957, pp. 172 (paper). $3.75.

HAROLD PALMER. *Psychopathic Personalities.* New York: Philosophical Library, 1957, pp. 179. $4.75.

E. LAKIN PHILLIPS AND JAMES F. GIBSON. *Psychology and Personality.* Englewood Cliffs, New Jersey: Prentice-Hall, 1957, pp. 338. $4.95.

SIDNEY L. PRESSY AND RAYMOND G. KUHLEN. *Psychological Development Through the Life Span.* New York: Harper & Brothers, 1957, pp. 654. $6.00.

Regional Problems and Issues in Human Relations Education. New York: Anti-Defamation League of B'nai B'rith, pp. 78 (paper).

HERBERT A. SIMON. *Models of Man: Mathematical Essays on Rational Behavior in a Social Setting.* New York: John Wiley & Sons, Inc., 1957, pp. 287. $5.00.

D. H. STOTT. *Unsettled Children and Their Families.* New York: Philosophical Library, 1957, pp. 240. $6.00.

W. L. SUMNER. *Visual Methods in Education.* Second Edition. New York: Philosophical Library, 1957, pp. 231. $6.00.

PERCIVAL M. SYMONDS. *Dynamics of Psychotherapy: The Psychology of Personality Change. Volume II. Process.* New York: Grune & Stratton, 1957, pp. 398, $6.50.

MILDRED C. TEMPLIN. *Certain Language Skills in Children: Their Development and Interrelationships.* Minneapolis: University of Minnesota Press, 1957, pp. 183. $4.00.

ARTHUR E. TRAXLER. *Techniques of Guidance.* Revised Edition. New York: Harper & Brothers, 1957, pp. 374. $6.00.

CLARK E. VINCENT. *Readings in Marriage Counseling.* New York: Thomas Y. Crowell Co., 1957, pp. 500. $4.50.

CHARLES H. WILSON. *A Teacher Is a Person.* New York: Henry Holt and Co., 1956, pp. 285. $3.75.

ARTHUR WORMHOUDT. *Hamlet's Mouse Trap: A Psychoanalytical Study of the Drama.* New York: Philosophical Library, pp. 221. $3.50.

Psychology and Mental Health

A report of the Institute on Education and Training for Psychological Contributions to Mental Health, held at Stanford University in August, 1955.

Edited by CHARLES R. STROTHER

Topics discussed at the Institute:

Training Needs of Psychologists in Community Mental Health

Specialization in Training

Practicum Training

Training for Therapy

Training for Research in the Mental Health Field

Problems of Departmental Organization

Price: $1.75

Order from:

American Psychological Association, Inc.
1333 Sixteenth Street, N. W.
Washington 6, D. C.